PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/82, 15/29, 1/19, C12Q 1/68, A01H 5/00

A1

(11) International Publication Number: WO 98/53086

(43) International Publication Date: 26 November 1998 (26.11.98)

(21) International Application Number: PCT/US98/10465

(22) International Filing Date: 21 May 1998 (21.05.98)

(30) Priority Data:
60/047,568 22 May 1997 (22.05.97) US

(71) Applicant *(for all designated States except US)*: WASHINGTON STATE UNIVERSITY RESEARCH FOUNDATION [US/US]; 1615 Eastgate Boulevard, N.E., Pullman, WA 99163 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: GRIMES, Howard, D. [US/US]; 1200 Douglas Drive, N.W., Pullman, WA 99163 (US). CHAO, Wun, S. [US/US]; 458 Sunset Drive, N.W., Pullman, WA 99163 (US).

(74) Agent: EARP, David, J.; Klarquist, Sparkman, Campbell, Leigh & Whinston, LLP, Suite 1600, One World Trade Center, 121 Salmon Street, S.W., Portland, OR 97204 (US).

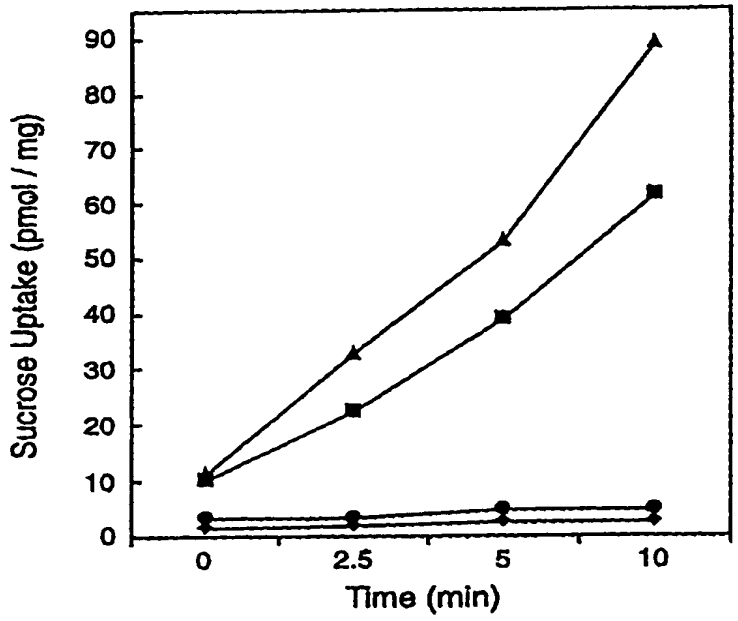
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SUCROSE-BINDING PROTEINS



(57) Abstract

A cDNA encoding a plant sucrose-binding protein (SBP) is provided, together with modified SBPs having enhanced sucrose uptake activity in a yeast assay system. Nucleic acid vectors, transgenic cells and transgenic plants having modified sucrose uptake activity are also provided. The invention also relates to promoter sequences useful for controlling expression of transgenes in plants, including transgenes.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

SUCROSE-BINDING PROTEINS

FIELD OF THE INVENTION

This invention relates to carbohydrate metabolism in plants, and in particular to sucrose-binding proteins (SBPs). Aspects of the invention include a novel SBP gene isolated from soybean, and modified SBPs having enhanced sucrose uptake activity. Nucleic acid vectors, transgenic cells and transgenic plants having modified sucrose uptake activity are also provided. The invention also relates to promoter sequences useful for controlling expression of transgenes in plants, including SBP transgenes.

BACKGROUND OF THE INVENTION

The regulation of sucrose transport in plants has a major impact on plant growth and productivity. Through photosynthesis, plants fix atmospheric carbon dioxide into triose phosphates, which are then used to produce sucrose and other carbohydrates. These carbohydrates are then transported throughout the plant for use as energy sources, carbon skeletons for biosynthesis and storage for future growth needs. Sucrose is the major form of transported carbohydrate. The ability of plant cells actively to transport sucrose across the plasma membrane so that the sucrose that is mobilized in the phloem can be taken into cells for use is a critical step in sucrose utilization.

The development of plant seeds involves the accumulation of carbon and nitrogen reserves in forms that can both withstand desiccation and be utilized as an energy source by the developing embryo during germination. The accumulation of carbon in developing seeds is mediated by specific plasma membrane proteins (Overvoorde et al., 1996; Riesmeier et al., 1992; Bush, 1993). Photoaffinity labeling of membranes isolated from soybean cotyledon tissue with a photolyzable sucrose analog identified a distinct 62 kD sucrose-binding protein, or SBP (Ripp et al., 1988). Analysis of the cDNA encoding the SBP and its deduced amino acid sequence indicates that the SBP contains a single hydrophobic domain at its N-terminus but otherwise is a hydrophilic protein lacking the expected membrane-spanning hydrophobic segments typically present in transport proteins (Grimes et

al., 1992). Biochemical analysis of the topology of the SBP demonstrates that it is tightly associated with the external leaflet of the plasma membrane (Overvorrde & Grimes, 1994). The involvement of the SBP in sucrose uptake was implicated by immunolocalization experiments demonstrating that the SBP is exclusively associated with the plasma membrane of cells involved in active sucrose uptake (Grimes et al., 1992). Kinetic analysis of SBPmediated sucrose uptake in a yeast system indicates that the uptake is specific for sucrose but is proton independent and relatively nonsaturable, thus defining a novel mechanism for sucrose uptake (Overvoorde et al., 1996).

Sucrose uptake in developing seeds affects two significant agricultural characteristics of the mature seed: the carbohydrate content of the resulting seed grain, and the vitality of the seedling that emerges when the seed grain is planted. Enhanced sucrose uptake activity in developing seeds may be desirable where it is an advantage to increase the carbohydrate content of the seed (e.g., where the seed is the primary plant material harvested, such as soybean). In contrast, decreased sucrose uptake activity in seeds might be desirable where the vegetative material of the plant is harvested. Thus, plants having modified sucrose uptake activity during seed development would be of significant agricultural importance, and it is to such plants that the present invention is directed.

SUMMARY OF THE INVENTION

The present invention provides isolated nucleic acid molecules encoding plant sucrose binding proteins, which are key proteins in the uptake of sucrose into developing seeds. In one embodiment, the invention provides modified forms of sucrose binding proteins that are shown to have enhanced sucrose uptake activity.

The previously described sucrose binding protein from soybean (Overvoode et al., 1996) is herein referred to as SBP1. A new SBP is provided herein and is referred to as SBP2. The SBP2 polypeptide is shown to be 489 amino acid residues in length, and to be expressed at enhanced levels during seed development. The SBP2 polypeptide is shown to have sucrose uptake activity in a heterologous yeast assay system.

In addition, modified forms of the SBP1 and SBP2 proteins are provided having enhanced sucrose uptake activity. In one embodiment, such forms are deletion mutants in which amino acid residues are removed from the C-terminus of the proteins. By way of example, removal of 80 amino acid residues from the C-terminus of the SBP1 protein is shown to produce increased sucrose uptake in the yeast assay system.

The invention also provides 5' regulatory regions (including promoter sequences) of the soybean *SBP1* and *SBP2* genes. These regulatory regions confer specific or enhanced expression in developing seeds and so may be used to express any transgene in developing seeds.

Thus, in one aspect, the invention provides a modified plant sucrose binding protein wherein the modified sucrose binding protein has a modified amino acid sequence compared to a corresponding wild-type sucrose binding protein, and wherein expression of the modified sucrose binding protein in a yeast assay system confers enhanced sucrose uptake compared to the corresponding wild-type sucrose binding protein. In particular embodiments, modified sucrose binding proteins provided by the invention enhance sucrose uptake in the yeast assay system by at least 10%, and preferably by at least 25%, compared to the wild-type sucrose binding protein. In certain embodiments, the modified plant sucrose binding proteins have a modified amino acid sequence comprising a C-terminal truncation compared to the wild-type sucrose binding protein. Such a truncation is typically of between about 10 and about 100 amino acids, and is preferably of about 80 amino acids. Although such modified SBPs may be produced from any known sucrose binding proteins, modified forms of SBP1 and SBP2 are exemplary of the invention. Modified forms of SBP1 and SBP2 include those forms having the amino acid sequences shown in Seq. I.D. Nos. 2 and 4, respectively.

In another aspect of the invention, nucleic acid molecules encoding modified plant sucrose binding proteins are provided, together with vectors comprising such nucleic acid molecules. The invention also provides transgenic plants expressing modified sucrose binding proteins. Such transgenic plants may have modified sucrose uptake activity, particularly in developing seeds.

In another aspect, the invention provides an isolated nucleic acid molecule encoding a SBP2 sucrose binding protein or a variant of a SBP2 protein. Such proteins may comprise an amino acid sequence as shown in Seq. I.D. Nos. 3 and 4, or sequences having at least 70% and preferably at least 90% sequence identity with these sequences. Recombinant expression cassettes comprising such nucleic acid molecules are also provided by the invention, as are transgenic plants comprising such recombinant expression cassettes.

Another aspect of the invention is a recombinant nucleic acid molecule comprising a promoter sequence operably linked to a nucleic acid sequence, wherein the promoter sequence comprises a *SBP1* or *SBP2* promoter. Such promoters preferably comprise at least 25 consecutive nucleotides of the 5' regulatory sequences shown in Seq. I.D. Nos. 6 and 7. In particular embodiments, the nucleic acid sequence comprises a plant sucrose binding protein. Transgenic plants comprising such recombinant nucleic acid molecules are also an aspect of the invention.

These and other aspects of the invention are discussed in more detail in the following description.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows an alignment of the SBP1 and SBP2 protein sequences.

Fig. 2 is a graph showing sucrose uptake activity in the yeast assay system.

SEQUENCE LISTING

The nucleic and amino acid sequences listed in the sequence listing are shown using standard single-letter abbreviations for nucleotide bases, and three-letter code for amino acids. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood to be included by any reference to the displayed strand.

Seq. I.D. No. 1 shows the amino acid sequence of the SBP1 protein.

Seq. I.D. No. 2 shows the amino acid sequence of the truncated SBP1 protein from which the C-terminus 80 amino acids are deleted.

Seq. I.D. No. 3 shows the amino acid sequence of the SBP2 protein.

Seq. I.D. No. 4 shows the amino acid sequence of the truncated SBP2 protein from which the C-terminus 80 amino acids are deleted.

Seq. I.D. No. 5 shows the *SBP2* cDNA sequence.

Seq. I.D. No. 6 shows the *SBP2* gene 5' regulatory region.

Seq. I.D. No. 7 shows the *SBP1* gene 5' regulatory region.

Seq. I.D. Nos. 8-14 show oligonucleotides that may be used to amplify various regions of the *SBP2* cDNA or 5' regulatory region.

DETAILED DESCRIPTION OF THE INVENTION

I. Methods

Standard molecular biology methods may be used to practice the present invention. Such methods are described in many publications, including Sambrook et al., (1989), Ausubel et al. (1994), Innis et al. (1990), Weissbach & Weissbach (1989), Tijssen (1993).

II. Definitions

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, Genes V published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew et al (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8). The nomenclature for DNA bases as set forth at 37 CFR § 1.822 and the standard three letter codes for amino acid residues are used herein.

In order to facilitate review of the various embodiments of the invention, the following definitions of terms is provided:

Sucrose binding protein (SBP) SBPs are involved in sucrose uptake in plants. This activity can be conveniently determined and measured using the yeast sucrose uptake assay originally described by Overvoorde et al. (1996), which is also described in detail below; in this assay system, SBPs confer sucrose uptake ability

on yeast cells that are otherwise unable to take up sucrose. Use of the term SBP refers generally to any sucrose binding protein, including the sucrose binding protein previously described by Grimes et al. (1992). This invention provides a cDNA encoding a previously unreported sucrose binding protein, the SBP2 protein from soybean (*Glycine max*). However the invention is not limited to this particular SBP: other nucleotide sequences which encode SBP enzymes are also part of the invention, including variants on the disclosed *Glycine* gene sequences and orthologous sequences from other plant species, the cloning of which is now enabled. Such sequences share the essential functional characteristic of encoding an enzyme that is capable of mediating sucrose uptake in the described yeast assay system. Nucleic acid sequences that encode SBPs and the proteins encoded by such nucleic acids share not only this functional characteristic, but also a specified level of sequence similarity (or sequence identity), as addressed below. The concept of sequence identity can also be expressed in the ability of two sequences to hybridize to each other under stringent conditions.

The present invention also provides modified SBPs having altered functional characteristics, as well as nucleic acid sequences encoding such proteins. An SBP isolated from an untransformed (wild-type) plant may be referred to as having a wild-type amino acid sequence. Modified SBPs have amino acid sequences that differ from the wild-type amino acid sequence. Such differences may take the form of amino acid deletions, additions, substitutions or truncations. A protein having amino acid deletions lacks one or more of the amino acid residues present in the wild-type sequence; such residues may be deleted from any portion of the protein. In contrast, a truncated protein is one in which one or more amino acids are deleted from the N and/or C terminus of the protein. Thus, truncated proteins are a sub-class of proteins having amino acid deletions.

Nucleic acid sequences encoding modified SBPs can readily be produced using standard methodologies, such as site directed mutagenesis and polymerase chain reaction amplification.

Sequence identity: the similarity between two nucleic acid sequences, or two amino acid sequences is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured

in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar the two sequences are.

The calculation of percentage of sequence identity for amino acid sequences may take into account conservative amino acid substitutions. Conservative amino acid substitutions involve the replacement of one amino acid residue with another residue having similar chemical and biological properties (e.g., charge or hydrophobicity). Such substitutions typically do not change the functional properties of the protein, and should therefore be accounted for in the calculation of sequence identity by assigning a value that is in between values assigned for identity (i.e., no change at that amino acid position) and non-conservative residue changes. Thus, conservative amino acid changes are scored as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. For example, if an identical amino acid is given a score of one and a non-conservative substitution is given a score of zero, a conservative substitution might be given a score of 0.5. The scoring of conservative substitutions is calculated, *e.g.*, according to the algorithm of Meyers and Miller (1988) *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith and Waterman (1981); Needleman and Wunsch (1970); Pearson and Lipman (1988); Higgins and Sharp (1988); Higgins and Sharp (1989); Corpet et al. (1988); Huang et al. (1992); and Pearson et al. (1994). Altschul et al. (1994) presents a detailed consideration of sequence alignment methods and homology calculations.

The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) is available from several sources, including the National Center for Biological Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. It can be accessed at <http://www.ncbi.nlm.nih.gov/BLAST/>. A description of how to determine sequence identity using this program is available at <http://www.ncbi.nlm.nih.gov/BLAST/blast_help.html>

Homologs of the disclosed SBP2 protein are characterized by possession of at least 80% sequence identity counted over the full length alignme

disclosed amino acid sequence of the soybean SBP2 amino acid sequence using the NCBI Blast 2.0, gapped blastp set to default parameters. Such homologous peptides will more preferably possess at least 85%, more preferably at least 90% and still more preferably at least 95% sequence identity determined by this method. When less than the entire sequence is being compared for sequence identity, homologs will possess at least 90% and more preferably at least 95% and more preferably still at least 98% sequence identity over short windows of 10-20 amino acids. Methods for determining sequence identity over such short windows are described at <http://www.ncbi.nlm.nih.gov/BLAST/blast_FAQs.html>. Homologs having the sequence identities described above will also possess the ability to mediate sucrose uptake in the described yeast assay system. The present invention provides not only the peptide homologs are described above, but also nucleic acid molecules that encode such homologs.

Homologs of the soybean *SBP2* gene are similarly characterized by possession of at least 70% sequence identity counted over the full length alignment with the disclosed *Glycine* SBP2 gene sequence using the NCBI Blast 2.0, gapped blastn set to default parameters. Such homologous nucleic acids will more preferably possess at least 75%, more preferably at least 80% and still more preferably at least 90% or 95% sequence identity determined by this method. When less than the entire sequence is being compared for sequence identity, homologs will possess at least 85% and more preferably at least 90% and more preferably still at least 95% sequence identity over 30 nucleotide windows. Homologs having the sequence identities described above will, in some embodiments, also encode a polypeptide having ability to mediate sucrose uptake in the described yeast assay system. However, homologs as defined above are useful for modifying sucrose uptake activity in transgenic plants (for example, as used in antisense constructs) even when they do not encode a functional peptide.

Another indication that two nucleic acid molecules are substantially homologous is that the two molecules hybridize to each other under stringent conditions when one molecule is used as a hybridization probe, and the other is present in a biological sample, e.g., genomic material from a cell. Specific hybridization means that the molecules hybridize substantially only to each other

and not to other molecules that may be present in the genomic material. Stringent conditions are sequence dependent and are different under different environmental parameters. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Conditions for nucleic acid hybridization and calculation of stringencies can be found in Sambrook et al. (1989) and Tijssen (1993). Hybridization conditions and stringencies are further discussed below.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences, due to the degeneracy of the genetic code. It is understood that changes in nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequence that all encode substantially the same protein.

Probes and primers: Nucleic acid probes and primers may readily be prepared based on the nucleic acids provided by this invention. A probe comprises an isolated nucleic acid attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, e.g., in Sambrook et. al. (1989) and Ausubel et al. (1987).

Primers are short nucleic acids, preferably DNA oligonucleotides 15 nucleotides or more in length. Primers may be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR), or other nucleic-acid amplification methods known in the art.

Methods for preparing and using probes and primers are described, for example, in Sambrook et al. (1989), Ausubel et al. (1987), and Innis et al., (1990). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 199 Whitehead Institute for Biomedical Research, Cambridge, MA). One of skill in the

art will appreciate that the specificity of a particular probe or primer increases with its length. Thus, for example, a primer comprising 20 consecutive nucleotides of the *SBP1* or *SBP2* gene 5' regulatory region will anneal to a target sequence (e.g., a corresponding SBP regulatory region from Faba bean) with a higher specificity than a corresponding primer of only 15 nucleotides. Thus, in order to obtain greater specificity, probes and primers may be selected that comprise 20, 25, 30, 35, 40, 50 or more consecutive nucleotides of the nucleic acid sequences disclosed herein.

Transformed: A transformed cell is a cell into which has been introduced a nucleic acid molecule by molecular biology techniques. As used herein, the term transformation encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including *Agrobacterium* transformation, plasmid transformation, viral tranafection and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

Vector: A nucleic acid molecule as introduced into a host cell, thereby producing a transformed host cell. A vector may include nucleic acid sequences that permit it to replicate in the host cell, such as an origin of replication. A vector may also include one or more selectable marker genes and other genetic elements known in the art.

Isolated: An "isolated" biological component (such as a nucleic acid or protein) has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, i.e., other chromosomal and extrachromosomal DNA and RNA, and proteins. Nucleic acids and proteins which have been "isolated" thus include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

Purified: The term purified does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified SBP preparation is one in which the SBP is more enriched than the protein is in its natural environment within a cell. Preferably, a preparation of SBP is purified such that the SBP represents at least 50% of the total protein content of the preparation.

Operably linked: A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein coding regions, in the same reading frame.

Recombinant: A recombinant nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques.

Ortholog: Two nucleotide or amino acid sequences are orthologs of each other if they share a common ancestral sequence and diverged when a species carrying that ancestral sequence split into two species. Orthologous sequences are also homologous sequences.

Transgenic plant: as used herein, this term refers to a plant that contains recombinant genetic material not normally found in plants of this type and which has been introduced into the plant in question (or into progenitors of the plant) by human manipulation. Thus, a plant that is grown from a plant cell into which recombinant DNA is introduced by transformation is a transgenic plant, as are all offspring of that plant which contain the introduced DNA (whether produced sexually or asexually). Transgenic plants may be produced from any transformable plant species, both monocotolydenous and dicotyledenous plants, including but not limited to soybean, rice, wheat, barley, and maize.

III. The *SBP2* cDNA and encoded SBP2 peptide

The nucleic acid sequence of the SBP2 cDNA is shown in Seq. I.D. No. 5, and the amino acid sequence of the SBP2 protein is shown in Seq. I.D. No. 3. A comparison of the amino acid sequences of SBP1 and SBP2 is shown in Fig. 1.

i. Differential expression of *SBP1* and *SBP2* genes in soybean leaves and cotyledons.

The sense and antisense RNAs of ^{32}P-labeled *SBP1* and *SBP2* 5'-flanking region were synthesized in vitro and 5.3 x 10^5 cpm of a *SBP1* sense, *SBP1* antisense, *SBP2* sense or *SBP2* antisense RNA probe were hybridized with 5 μg poly(A+) mRNA from soybean leaves and cotyledons. *SBP1* and *SBP2* transcripts were observed to accumulate to similar levels in soybean cotyledons. In contrast, no *SBP1* and *SBP2* transcripts were detected in 4-wk old soybean leaves.

ii. Differential Expression of Soybean *SBP1* and *SBP2* genes

The expression patterns of the *SBP1* and *SBP2* genes were examined in soybean seeds using RNase protection methods. Five stages of seed cotyledon development were used (Stage 1 = or < 4 mm, Stage 2 = 5-6 mm, Stage 3 = 7 mm, Stage 4 = 9 mm, Stage 5 = 11-12 mm). During cotyledon development, an *SBP1* antisense probe protected three major fragment (119, 111, and 97 nucleotides), indicating that three different transcription start sites were used. The *SBP1* mRNA level reaches a plateau at stage 3, and this expression level is maintained until stage 5. In contrast, 5 protected fragments were detected when using *SBP2* antisense probe, and *SBP2* mRNA level continuously increased until seed size reached 11-12 mm. Quantitative data indicated that *SBP1* mRNA level is three time more abundant than that of *SBP2*. The mRNA level of leaf tip is very low. However, low levels of *SBP1* mRNA can be observed in 3 mm leaf tips after prolonged exposure. These data indicate that both *SBP1* and *SBP2* mRNAs are actively and differentially transcribed during seed development.

IV. 5' regulatory regions of *SBP1* and *SBP2*

Given the tissue-specific expression of the *SBP1* and SBP2 genes, the regulatory regions of these genes responsible for conferring such expression are of interest, and may be used to regulate transgene expression in a similarly tissue-specific manner.

The 5' regulatory regions of SBP1 and SBP2 are shown in Seq. I.D. Nos. 6 and 7, respectively.

V. Modified SBPs having enhanced sucrose uptake activity

The yeast assay system described by Overvoorde et al (1996) was used to determine the effect of modifying the amino acid sequence of the SBP proteins. This assay uses a derivative of the yeast strain *susy7* (Riesmeier et al., 1992) which has a spinach sucrose synthase cDNA stably integrated into its genome to mediate the intracellular hydrolysis of sucrose. However, this yeast strain lacks the ability to transport sucrose and so is unable to grow on a medium containing sucrose as the sole carbon source (Riesmeier et al., 1992). To generate a host strain that permits selection for yeast transformed with a sucrose binding protein gene, the *susy7* strain was selected for uracil auxotrophy by growth on medium containing 5'-fluoroorotic acid (Overvoorde et al., 1996). The resulting strain, *susy7/ura3* is unable to grow on a medium lacking uracil and containing glucose as the sole carbon source.

Chimeric genes consisting of the yeast alcohol dehydrogenase1 (*ADH1*) promoter, an SBP open reading frame and the *ADH1* polyadenylation signal were constructed in the yeast vector pMK195 as described by Overvoorde et al. (1996) to create plasmids designated pYESBP. The *susy7/ura3* yeast strain was transformed with these constructs using a small-scale LiOAc-based procedure essentially as described by Gietz et al. (1992). Transformed yeast were then plated on the uracil dropout selection medium containing 2% glucose (CM[GLU]) or 2% sucrose (CM[SUC]) (Ausubel et al., 1994).

Uptake assays were performed by growing the transformed yeast cells to an OD_{600} of 0.5 to 1.3 in YPD, harvested by centrifugation, washed twice with 25 mM Mes-KOH, pH 5.5, 0.5 – 2.5 µCi of ^{14}C sucrose, and unlabeled sucrose at twice the final concentration. Aliquots of the uptake solution and cells were collected at specified time points, and uptake was quenched by transfer to 5 ml of ice-cold water. The cells were collected by filtration through glass fiber filters and washed five times with 5 ml of ice-cold water. The radioactivity taken up by the cells was determined by liquid scintillation counting. All uptake assays were performed in a final concentration of 1 mM sucrose.

Nucleic acid sequences encoding modified forms of the SBP1 protein were constructed and introduced into the pYESBP constructs described above. Fig. 2 shows the sucrose uptake rate obtained with yeast cells transformed with the pMK195 vector only (filed circles), and constructs expressing the full length SBP1 protein (filled square) and a truncated SBP1 protein missing the C-terminal 80 amino acids (filled triangle). The amino acid sequence of this truncated SBP1 protein is shown in Seq. I.D. No. 2. The truncated protein comprises residues 1-444 of the full length SBP1.

This surprising result indicates that enhanced sucrose uptake in plants may be achieved by introducing transgenes encoding modified SBPs. Modified SBPs having enhanced sucrose uptake activity include forms of SBP1 and SBP2 having C-terminal deletions. Such deletions include removal of about 80 amino acids from the C-terminal, but deletions of greater or fewer than 80 amino acids may also be employed. The sucrose uptake activity any particular deletion may readily be determined using the yeast sucrose uptake assay described above. Thus, by way of example, SBP proteins having C-terminal deletions of between 10 and 100 amino acids are candidates for enhanced sucrose uptake activity and may be assayed using this system.

EXAMPLES

The following examples are illustrative of various embodiments of the present invention.

Example one: Preferred method for producing SBP nucleic acids

This invention provides a *SBP2* cDNA sequence and the amino acid sequence of the SBP2 protein, modified SBP proteins having enhanced sucrose uptake activity, and 5' regulatory regions for the *SBP1* and *SBP2* genes. The polymerase chain reaction (PCR) may now be utilized in a preferred method for producing nucleic acid sequences encoding the various SBP proteins described in the invention, as well as the *SBP* gene 5' regulatory regions. PCR amplification of cDNAs encoding the SBP proteins of the present invention may be accomplished either by direct PCR from a plant cDNA library or by Reverse-Transcription PCR

(RT-PCR) using RNA extracted from plant cells as a template. Amplification of *SBP* gene sequences and 5' regulatory regions may be accomplished by direct PCR amplification from plant genomic DNA, or from a plant genomic library. Methods and conditions for both direct PCR and RTPCR are known in the art and are described in Innis et al. (1990).

The selection of PCR primers will be made according to the portions of the cDNA or gene that are to be amplified. Primers may be chosen to amplify small segments of the cDNA, the open reading frame, the entire cDNA molecule or the entire gene sequence. Variations in amplification conditions may be required to accommodate primers of differing lengths; such considerations are well known in the art and are discussed in Innis et al. (1990), Sambrook et al. (1989), and Ausubel et al (1992). By way of example only, the entire *SBP2* cDNA molecule as shown in Seq. I.D. No. 5 may be amplified using the following combination of primers:

primer 1 5' TGTAAAACGACGGCCAGTGAATT 3' (Seq. I.D. No. 8)

primer 2 5' GATTACGCCAAGCTCGAAATTAA 3' (Seq. I.D. No. 9)

The open reading frame portion of the *SBP2* cDNA may be amplified using the following primer pair:

primer 3 5' ATGGCGACCAGAGCCAAGCTTTCTTTA 3' (Seq. I.D. No 10)

primer 4 5' CGCAACAGCGCGACGACCACGCTCGCT 3' (Seq. I.D. No. 11)

And a cDNA encoding a truncated version of the SBP2 protein (having the C-terminal 80 amino acids removed) may be amplified using the following primer pair:

primer 3 5' ATGGCGACCAGAGCCAAGCTTTCTTTA 3' (Seq. I.D. No. 10)

primer 5 5' GAAGGGATGACCAGGAGGGACAACAAA 3' (Seq. I.D. No. 12)

The SBP2 5 regulatory sequence may be amplified using the following primer pair:

primer 6 5' TTGTAAACGACGGCCAGTGAATT 3' (S[illegible]

primer 7 5' GGTGAGGTCAGTGAGGAACAACA 3' (Seq. I.D. No. 14)

These primers are illustrative only; it will be appreciated by one skilled in the art that many different primers may be derived from the provided nucleic acid sequences in order to amplify particular regions of these molecule. Resequencing of PCR products obtained by these amplification procedures is recommended; this will facilitate confirmation of the amplified sequence and will also provide information on natural variation on this sequence in different ecotypes and plant populations.

Oligonucleotides that are derived from the *SBP2* cDNA or *SBP1* and *SBP2* 5' regulatory regions are encompassed within the scope of the present invention. Preferably, such oligonucleotide primers will comprise a sequence of at least 15-20 consecutive nucleotides of the *SBP2* cDNA or gene sequences. To enhance amplification specificity, oligonucleotide primers comprising at least 25, 30, 35, 40, 45 or 50 consecutive nucleotides of these sequences may also be used.

In addition, the *SBP2* gene sequence may be obtained by PCR amplification using primers derived from the disclosed cDNA sequence to probe a genomic library or genomic DNA, or by probing a genomic DNA library using a labeled probe derived from the *SBP2* cDNA sequence. Standard PCR amplification or hybridization methods may be used for these approaches.

Example Two : Isolation of homologous gene sequence from other plant species

With the provision herein of the soybean *SBP2* cDNA, SBP 5' regulatory regions, and the disclosed discovery that modification of SBP proteins, particularly truncation of the C-terminus, produces enhanced sucrose uptake, the invention also enables the production of corresponding molecules from other plant species. Thus, the present invention permits the isolation of *SBP2* homologs from other species, as well as the production of enhanced efficiency SBP proteins of other plant species. Both conventional hybridization and PCR amplification procedures may be utilized to obtain corresponding cDNAs from other species and to produce nucleic acids encoding enhanced activity SBP proteins. Common to both of these techniques is the hybridization of probes or primers derived from the *SBP2* cDNA or gene

sequence to a target nucleotide preparation, which may be, in the case of conventional hybridization approaches, a cDNA or genomic library or, in the case of PCR amplification, a cDNA or genomic library, or an mRNA preparation.

Direct PCR amplification may be performed on cDNA or genomic libraries prepared from the plant species in question, or RT-PCR may be performed using mRNA extracted from the plant cells using standard methods. PCR primers will comprise at least 15 consecutive nucleotides of the *SBP2* cDNA. One of skill in the art will appreciate that sequence differences between the soybean *SBP2* cDNA and the target nucleic acid to be amplified may result in lower amplification efficiencies. To compensate for this, longer PCR primers or lower annealing temperatures may be used during the amplification cycle. Where lower annealing temperatures are used, sequential rounds of amplification using nested primer pairs may be necessary to enhance specificity.

For conventional hybridization, the hybridization probe is preferably conjugated with a detectable label such as a radioactive label, and the probe is preferably of at least 20 nucleotides in length. As is well known in the art, increasing the length of hybridization probes tends to give enhanced specificity. The labeled probe derived from the soybean *SBP2* cDNA or gene sequence may be hybridized to a plant cDNA or genomic library and the hybridization signal detected using means known in the art. The hybridizing colony or plaque (depending on the type of library used) is then purified and the cloned sequence contained in that colony or plaque isolated and characterized.

Homologs of the soybean *SBP2* cDNA may alternatively be obtained by immunoscreening of an expression library. With the provision herein of the disclosed *SBP2* nucleic acid sequences, the enzyme may be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the SBP2 protein. Antibodies may also be raised against synthetic peptides derived from the SBP2 amino acid sequence presented herein. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988). Such antibodies can then be used to screen an expression cDNA library produced from the plant from which it is desired to

clone the *SBP2* ortholog, using the methods described above. The selected cDNAs can be confirmed by sequencing and enzyme activity.

The soybean *SBP2* gene or cDNA, and homologs of these sequences from other plants may be incorporated into transformation vectors and introduced into plants to modify SBP activity in such plants, as described in Example Three below. In addition, nucleic acids encoding modified SBP proteins as taught herein may also be used to produce plants having modified sucrose uptake activity. It is anticipated that the native SBP gene promoter may be particularly useful in the practice of the present invention in that it may be used to drive the expression of SBP transgenes, such as antisense constructs. By using the native SBP gene promoter, expression of these transgenes may be regulated in coordination with the native SBP gene (for example, in the same temporal or tissue-specific expression patterns).

Example Three: Transgenic plants having modified sucrose uptake activity

Once a gene (or cDNA) encoding a protein involved in the determination of a particular plant characteristic has been isolated, standard techniques may be used to express the cDNA in transgenic plants in order to modify that particular plant characteristic. The basic approach is to clone the cDNA into a transformation vector, such that it is operably linked to control sequences (e.g., a promoter) that direct expression of the cDNA in plant cells. The transformation vector is then introduced into plant cells by one of a number of techniques (e.g., electroporation) and progeny plants containing the introduced cDNA are selected. Preferably all or part of the transformation vector will stably integrate into the genome of the plant cell. That part of the transformation vector which integrates into the plant cell and which contains the introduced cDNA and associated sequences for controlling expression (the introduced "transgene") may be referred to as the recombinant expression cassette.

Selection of progeny plants containing the introduced transgene may be made based upon the detection of an altered phenotype. Such a phenotype may result directly from the cDNA cloned into the transformation vector or may be manifested as enhanced resistance to a chemical agent (such as an antibiotic) as a result of the

inclusion of a dominant selectable marker gene incorporated into the transformation vector.

The choice of (a) control sequences and (b) how the cDNA (or selected portions of the cDNA) are arranged in the transformation vector relative to the control sequences determine, in part, how the plant characteristic affected by the introduced cDNA is modified. For example, the control sequences may be tissue specific, such that the cDNA is only expressed in particular tissues of the plant (e.g., pollen, seed) and so the affected characteristic will be modified only in those tissues. The cDNA sequence may be arranged relative to the control sequence such that the cDNA transcript is expressed normally, or in an antisense orientation. Expression of an antisense RNA corresponding to the cloned cDNA will result in a reduction of the targeted gene product (the targeted gene product being the protein encoded by the plant gene from which the introduced cDNA was derived). Over-expression of the introduced cDNA, resulting from a plus-sense orientation of the cDNA relative to the control sequences in the vector, may lead to an increase in the level of the gene product, or may result in co-suppression (also termed "sense suppression") of that gene product.

Successful examples of the modification of plant characteristics by transformation with cloned cDNA sequences are replete in the technical and scientific literature. Selected examples, which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include:

U.S. Patent No. 5,451,514 to Boudet (modification of lignin synthesis using antisense RNA and co-suppression);

U.S. Patent No. 5,443,974 to Hitz (modification of saturated and unsaturated fatty acid levels using antisense RNA and co-suppression);

U.S. Patent No. 5,530,192 to Murase (modification of amino acid and fatty acid composition using antisense RNA);

U.S. Patent No. 5,455,167 to Voelker (modification of medium chain fatty acids)

U.S. Patent No. 5,231,020 to Jorgensen (modification of flavor [illegible] co-suppression);

U.S. Patent No. 5,583,021 to Dougherty (modification of virus resistance by expression of plus-sense untranslatable RNA);

WO 96/13582 (modification of seed VLCFA composition using over expression, co-suppression and antisense RNA in conjunction with the Arabidopsis FAE1 gene); and

WO 95/15387 (modification of seed VLCFA composition using over expression of jojoba wax synthesis gene).

These examples include descriptions of transformation vector selection, transformation techniques and the construction of constructs designed to over-express the introduced cDNA or to express antisense RNA corresponding to the cDNA. In light of the foregoing and the provision herein of the SBP2 gene and nucleic acids encoding modified SBP proteins conferring enhanced sucrose uptake activity, it is thus apparent that one of skill in the art will be able to introduce these nucleic acids, or homologous or derivative forms of these molecules (e.g., antisense forms), into plants in order to produce plants having modified sucrose uptake activity activity, in developing seeds and other tissues. The result can be altered plant development with agricultural and economic consequences.

a. Plant Types

Nucleic acid molecules according to the present invention (e.g., the SBP2 gene, nucleic acids encoding modified SBP proteins, homologs of these sequences and derivatives such as antisense forms) may be introduced into any plant type in order to modify sucrose uptake activity in the plant. Thus, the sequences of the present invention may be used to modify sucrose uptake activity in any higher plant, including monocotyledonous and dicotyledenous plants, including, but not limited to maize, wheat, rice, barley, soybean, beans in general, rape/canola, alfalfa, flax, sunflower, safflower, brassica, cotton, flax, peanut, clover; vegetables such as lettuce, tomato, cucurbits, potato, carrot, radish, pea, lentils, cabbage, broccoli, brussel sprouts, peppers; tree fruits such as apples, pears, peaches, apricots; flowers such as carnations and roses.

b. Vector Construction, Choice of Promoters

A number of recombinant vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants have been described including those described in Pouwels et al., (1987), Weissbach and Weissbach, (1989), and Gelvin et al., (1990). Typically, plant transformation vectors include one or more cloned plant genes (or cDNAs) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter regulatory region (*e.g.*, a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which may be useful for expressing nucleic acids include: the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in most plant tissues (*see, e.g.*, Odel et al., 1985, Dekeyser et al., 1990, Terada and Shimamoto, 1990; Benfey and Chua. 1990); the nopaline synthase promoter (An et al., 1988); and the octopine synthase promoter (Fromm et al., 1989).

A variety of plant gene promoters that are regulated in response to environmental, hormonal, chemical, and/or developmental signals, also can be used for expression of the cDNA in plant cells, including promoters regulated by: (a) heat (Callis et al., 1988; Ainley, et al. 1993; Gilmartin et al. 1992); (b) light (*e.g.*, the pea rbcS-3A promoter, Kuhlemeier et al., 1989, and the maize rbcS promoter, Schaffner and Sheen, 1991); (c) hormones, such as abscisic acid (Marcotte et al., 1989); (d) wounding (*e.g.*, *wunI*, Siebertz et al., 1989); and (e) chemicals such as methyl jasminate or salicylic acid (*see also* Gatz et al., 1997) can also be used to regulate gene expression.

Alternatively, tissue specific (root, leaf, flower, and seed for example) promoters (Carpenter et al., 1992; Denis et al., 1993; Opperman et al., 1993; Stockhause et al. 1997; Roshal et al., 1987; Schernthaner et al., 1988; and Bustos et al., 1989) can be fused to the coding sequence to obtained particular expression in respective organs. In addition, the timing of the expression can be controlled by

using promoters such as those acting at senescencing (Gan and Amasino, 1995) or late seed development (Odell et al., 1994).

The promoter regions of the *SBP1* and *SBP2* genes disclosed herein confer developing seed-specific expression in soybean. Accordingly, these promoters may be used to obtain developing seed specific expression of the introduced transgene.

Plant transformation vectors may also include RNA processing signals, for example, introns, which may be positioned upstream or downstream of the ORF sequence in the transgene. In addition, the expression vectors may also include additional regulatory sequences from the 3'-untranslated region of plant genes, *e.g.*, a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Finally, as noted above, plant transformation vectors may also include dominant selectable marker genes to allow for the ready selection of transformants. Such genes include those encoding antibiotic resistance genes (*e.g.*, resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin or spectinomycin) and herbicide resistance genes (*e.g.*, phosphinothricin acetyltransferase).

c. Arrangement of *SBP* sequence in vector

The particular arrangement of the *SBP* sequence in the transformation vector will be selected according to the type of expression of the sequence that is desired.

Where enhanced sucrose uptake activity is desired in the plant, the SBP ORF may be operably linked to a constitutive high-level promoter such as the CaMV 35S promoter. Modification of sucrose uptake activity may also be achieved by introducing into a plant a transformation vector containing a variant form of the *SBP2* gene, for example a form which varies from the exact nucleotide sequence of the *SBP2* ORF, but which encodes a protein that retains the functional characteristic of the SBP2 protein, i.e., conferring sucrose uptake activity. By way of example, enhanced sucrose uptake activity may also be obtained by utilizing a nucleic acid sequence encoding a modified SBP as discussed above. Such modified SBPs include SBPs having C-terminal deletions, generally in the range of 10-100 amino acid residue, and preferably about 80 amino acid residues.

In contrast, a reduction sucrose uptake activity in the transgenic plant may be obtained by introducing into plants antisense constructs based on a SBP gene sequence. For antisense suppression, SBP gene is arranged in reverse orientation relative to the promoter sequence in the transformation vector. The introduced sequence need not be the full length SBP gene, and need not be exactly homologous to the SBP gene found in the plant type to be transformed. Generally, however, where the introduced sequence is of shorter length, a higher degree of homology to the native SBP sequence will be needed for effective antisense suppression. Preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides. Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous SBP gene in the plant cell. Although the exact mechanism by which antisense RNA molecules interfere with gene expression has not been elucidated, it is believed that antisense RNA molecules bind to the endogenous mRNA molecules and thereby inhibit translation of the endogenous mRNA.

Suppression of endogenous SBP gene expression can also be achieved using ribozymes. Ribozymes are synthetic RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 to Cech and U.S. Patent No. 5,543,508 to Haselhoff. The inclusion of ribozyme sequences within antisense RNAs may be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that bind to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Constructs in which a SBP nucleic acid (or variants thereof) are over-expressed may also be used to obtain co-suppression of the endogenous SBP gene in the manner described in U.S. Patent No. 5,231,021 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the SBP gene be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous SBP gene. However, as with antisense

suppression, the suppressive efficiency will be enhanced as (1) the introduced sequence is lengthened and (2) the sequence similarity between the introduced sequence and the endogenous SBP gene is increased.

Constructs expressing an untranslatable form of a SBP mRNA may also be used to suppress the expression of endogenous SBP activity. Methods for producing such constructs are described in U.S. Patent No. 5,583,021 to Dougherty et al. Preferably, such constructs are made by introducing a premature stop codon into the SBP ORF.

Finally, dominant negative mutant forms of the disclosed sequences may be used to block endogenous SBP activity. Such mutants require the production of mutated forms of the SBP protein that bind to sucrose but do not catalyze the uptake of sucrose.

d. Transformation and Regeneration Techniques

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells is now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods may include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumeficiens* (AT) mediated transformation. Typical procedures for transforming and regenerating plants are described in the patent documents listed at the beginning of this section.

e. Selection of Transformed Plants

Following transformation and regeneration of plants with the transformation vector, transformed plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic resistance on the seedlings of transformed plants, and selection of

transformants can be accomplished by exposing the seedlings to appropriate concentrations of the antibiotic.

After transformed plants are selected and grown to maturity, they can be assayed using known methods to determine whether SBP activity has been altered as a result of the introduced transgene. In addition, antisense or sense suppression of an endogenous SBP gene may be detected by analyzing mRNA expression on Northern blots.

Example Four: Production of sequence variants

As noted above, modification of sucrose uptake activity in plant cells can be achieved by transforming plants with the *SBP2* cDNA or gene, antisense constructs based on the *SBP2* cDNA or gene sequence or nucleic acid sequences encoding modified SBP proteins. With the provision of the *SBP2* cDNA and gene sequences and the SBP 5' regulatory regions herein, the creation of variants on these sequences by standard mutagenesis techniques is now enabled.

Variant DNA molecules include those created by standard DNA mutagenesis techniques, for example, M13 primer mutagenesis. Details of these techniques are provided in Sambrook et al. (1989), Ch. 15. By the use of such techniques, variants may be created which differ in minor ways from the disclosed sequences disclosed. DNA molecules and nucleotide sequences which are derivatives of those specifically disclosed herein and which differ from those disclosed by the deletion, addition or substitution of nucleotides while still encoding a protein which possesses the functional characteristic of a SBP protein (i.e., the ability to mediate sucrose uptake in the yeast assay system) are comprehended by this invention. DNA molecules and nucleotide sequences which are derived from the *SBP2* cDNA and gene sequences disclosed include DNA sequences which hybridize under stringent conditions to the DNA sequences disclosed, or fragments thereof.

Hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the hybridization method of choice and the composition and length of the hybridizing DNA used. Generally, the temperature of hybridization and the ionic strength (especially the Na^+ concentration) of the hybridization buffer will determine the stringency of hybridization. Calculations

regarding hybridization conditions required for attaining particular degrees of stringency are discussed by Sambrook et al. (1989), chapters 9 and 11, herein incorporated by reference. By way of illustration only, a hybridization experiment may be performed by hybridization of a DNA molecule (for example, soybean *SBP2* cDNA sequence) to a target DNA molecule (for example, the a corresponding *SBP2* cDNA sequence in tobacco) which has been electrophoresed in an agarose gel and transferred to a nitrocellulose membrane by Southern blotting (Southern, 1975), a technique well known in the art and described in (Sambrook et al., 1989). Hybridization with a target probe labeled with [^{32}P]dCTP is generally carried out in a solution of high ionic strength such as 6xSSC at a temperature that is 20-25° C below the melting temperature, T_m, described below. For such Southern hybridization experiments where the target DNA molecule on the Southern blot contains 10 ng of DNA or more, hybridization is typically carried out for 68 hours using 12 ng/ml radiolabeled probe (of specific activity equal to 10^9 CPM/μg or greater). Following hybridization, the nitrocellulose filter is washed to remove background hybridization. The washing conditions should be as stringent as possible to remove background hybridization but to retain a specific hybridization signal. The term T_m represents the temperature above which, under the prevailing ionic conditions, the radiolabeled probe molecule will not hybridize to its target DNA molecule. The T_m of such a hybrid molecule may be estimated from the following equation (Bolton and McCarthy, 1962):

$$T_m = 81.5\ C\ \ 16.6(\log_{10}[Na^+]) + 0.41(\%G+C)-0.63(\%\ \text{formamide})\ \ (600/l)$$

Where l = the length of the hybrid in base pairs.

This equation is valid for concentrations of Na^+ in the range of 0.01 M to 0.4 M, and it is less accurate for calculations of T_m in solutions of higher [Na^+]. The equation is also primarily valid for DNAs whose G+C content is in the range of 30% to 75%, and it applies to hybrids greater than 100 nucleotides in length (the behavior of oligonucleotide probes is described in detail in Ch. 11 of Sambrook et al., 1989).

Thus, by way of example, for a 150 base pair DNA probe derived from the first 150 base pairs of the open reading frame of the soybean SBP2 cDNA (with a

hypothetical %GC = 45%), a calculation of hybridization conditions required to give particular stringencies may be made as follows:

For this example, it is assumed that the filter will be washed in 0.3 xSSC solution following hybridization, thereby $[Na^+] = 0.045M$; %GC = 45%; Formamide concentration = 0; $l = 150$ base pairs; and $T_m = 81.5\ \ 16(\log_{10}[Na^+]) + (0.41 \times 45)\ (600/150)$ and so $T_m = 74.4$ C.

The T_m of double-stranded DNA decreases by 1-1.5 °C with every 1% decrease in homology (Bonner et al., 1973). Therefore, for this given example, washing the filter in 0.3 xSSC at 59.4-64.4 °C will produce a stringency of hybridization equivalent to 90%. Alternatively, washing the hybridized filter in 0.3 xSSC at a temperature of 65.4-68.4 °C will yield a hybridization stringency of 94%. The above example is given entirely by way of theoretical illustration. One skilled in the art will appreciate that other hybridization techniques may be utilized and that variations in experimental conditions will necessitate alternative calculations for stringency.

DNA sequences from plants that encode a protein having SBP activity and which hybridize under hybridization conditions of at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90% and most preferably at least 95% stringency to the disclosed *SBP2* sequence are encompassed within the present invention.

The degeneracy of the genetic code further widens the scope of the present invention as it enables major variations in the nucleotide sequence of a DNA molecule while maintaining the amino acid sequence of the encoded protein. For example, the second amino acid residue of the soybean SBP2 protein is alanine. This is encoded in the soybean SBP2 open reading frame by the nucleotide codon triplet GCG. Because of the degeneracy of the genetic code, three other nucleotide codon triplets-GCA, GCC and GCT-also code for alanine. Thus, the nucleotide sequence of the soybean SBP2 ORF could be changed at this position to any of these three codons without affecting the amino acid composition of the encoded protein or the characteristics of the protein. Based upon the degeneracy of the genetic code, variant DNA molecules may be derived from the cDNA and gene sequences disclosed herein using standard DNA mutagenesis techniques as

described above, or by synthesis of DNA sequences. Thus, this invention also encompasses nucleic acid sequences which encode a SBP protein but which vary from the disclosed nucleic acid sequences by virtue of the degeneracy of the genetic code.

The present invention teaches that enhanced sucrose uptake activity may be obtained by modifying the sequence of a plant SBP, e.g., by deleting 80 C-terminal amino acids. One skilled in the art will recognize that DNA mutagenesis techniques may be used not only to produce variant DNA molecules, but will also facilitate the production of such modified SBP protein. In addition, other changes to the amino acid sequence can be made including deletions, additions and substitutions.

While the site for introducing an amino acid sequence variation is predetermined, the mutation *per se* need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed protein variants screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence as described above are well known.

Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to more than 100 residues. Substitutions, deletions, insertions or any combination thereof may be combined to arrive at a final construct. Obviously, the mutations that are made in the DNA encoding the protein must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure.

Substitutional variants are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the following Table 1 when it is desired to finely modulate the characteristics of the protein. Table 1 shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative substitutions.

Table 1.

Original Residue	Conservative Substitutions
Ala	ser
Arg	lys
Asn	gln; his
Asp	glu
Cys	ser
Gln	asn
Glu	asp
Gly	pro
His	asn; gln
Ile	leu, val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

Substantial changes in enzymatic function or other features are made by selecting substitutions that are less conservative than those in Table 1, i.e., selecting residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histadyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

The effects of these amino acid substitutions or deletions or additions may be assessed for derivatives of the SBP proteins by analyzing the ability of the

derivative proteins to catalyze sucrose uptake in the yeast assay system described above.

Example Five: Use of SBP 5' regulatory regions to control transgene expression

The promoters of the *Glycine SBP1* and *SBP2* genes confer developing seed-specific expression. Accordingly, the promoter sequences, shown in Seq. I.D. Nos. 7 (*SBP2*) and 8 (*SBP1*) may be used to produce transgene constructs that are specifically expressed in developing seeds. One of skill in the art will recognize that regulation of transgene expression in developing seeds may be achieved with less than the entire 5' regulatory sequences shown in Seq. I.D. Nos. 7 & 8. Thus, by way of example, developing seed-specific expression may be obtained by employing a 50 base pair or 100 base pair region of the disclosed promoter sequences. The determination of whether a particular sub-region of the disclosed sequence operates to confer effective seed-specific expression in a particular system (taking into account the plant species into which the construct is being introduced, the level of expression required, etc.) will be performed using known methods, such as operably linking the promoter sub-region to a marker gene (e.g. GUS), introducing such constructs into plants and then determining the level of expression of the marker gene in developing seeds and other plant tissues.

The present invention therefore facilitates the production, by standard molecular biology techniques, of nucleic acid molecules comprising the *SBP1* or *SBP2* promoter sequence operably linked to a nucleic acid sequence, such as an open reading frame. Suitable open reading frames include open reading frames encoding any protein for which expression in developing seeds is desired. Examples of genes that may suitably be expressed in a seed-specific manner under the control of the disclosed SBP promoters include, but are not limited to:

(1) genes that enhance the nutritional quality of the seeds, for example, by increasing the content of limiting amino acids, including lysine, methionine and cysteine. This may be achieved by expressing proteins containing high levels of these amino acids in seeds. Examples include the high methionine storage proteins from brazil nut (Saalbach et al., 1996) and sunflower (Molvig et al., 1997).

(2) genes that increase gluten levels in wheat, so as to enhance the bread-making quality of the wheat flour (Shewry et al., 1995).

(3) genes that enhance insect resistance in the seed (for example, resistance to weevils). Suitable genes include the α-amylase inhibitor gene which kills seed weevils (Schmidt, 1994).

REFERENCES

Ainley et al. (1993) *Plant Mol. Biol.* **22**:13-23.

Altschul & Gish. (1996) *Methods Enzymol.*, **266**, 460-80.

Altschul et al. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.

Altschul et al. (1994) *Nature Genet.*, **6**, 119-29.

An et al. (1988) *Plant Physiol.* **88**:547.

Ausubel et al. (1987) In *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Intersciences.

Ausubel et al. eds., *Current Protocols in Molecular Biology*, John Wiley and Sons, New York, 1994.

Benfey & Chua (1990) *Science* **250**:959-966.

Birch et al. (1997) *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **48**:297-326.

Bolton & McCarthy (1962) *Proc. Natl. Acad. Sci. USA* **48**:1390.

Bonner et al. (1973) *J. Mol. Biol.* **81**:123.

Bush (1993) *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **44**:513542.

Bustos et al. (1989) *Plant Cell* **1**:839.

Callis et al. (1988) *Plant Physiol.* **88**:965.

Carpenter et al. (1992) *The Plant Cell* **4**:557-571.

Corpet et al. (1988) *Nucleic Acids Research* **16**, 10881-90

Dekeyser et al. (1990) *Plant Cell* **2**:591.

Denis et al. (1993) *Plant Physiol.* **101**:1295-1304.

Fromm et al. (1989) *Plant Cell* **1**:977.

Gan & Amansino (1995) *Science* **270**:1986-1988.

Gatz et al. (1997) *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **48**:89-108.

Gelvin et al. (1990) *Plant Molecular Biology Manual*, Kluwer Acad[illegible]

Gietz et al. (1992) *Nuc. Acids Res.* **20**:1425.

Gilmartin et al. (1992). *The Plant Cell* **4**:839-949.

Grimes et al. (1992) *Plant Cell* **4**:15611574.

Harlow & Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Laboratory, New York.

Higgins & Sharp (1988) *Gene,* **73**: 237-244

Higgins & Sharp (1989) *CABIOS* **5**: 151-153

Huang et al. (1992) *Computer Applications in the Biosciences* **8**, 155-65

Innis et al. (1990) *PCR Protocols, A Guide to Methods and Applications*, Innis et al. (eds.), Academic Press, Inc., San Diego, California.

Kawasaki et al. (1990). In *PCR Protocols, A Guide to Methods and Applications*, Innis et al. (eds.), 2127, Academic Press, Inc., San Diego, California.

Kuhlemeier et al. (1989) *Plant Cell* **1**: 471.

Meyer & Horgan (1996) *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **47**:23-48.

Meyers & Miller (1988) *Computer Applic. Biol. Sci.*, **4**: 11-17.

Molvig et al. (1997) *Proc. Natl. Acad. Sci.* **94**:8393-8398.

Needleman & Wunsch (1970) *J. Mol. Biol.* **48**: 443

Odel et al. (1985) *Nature* **313**:810.

Odell et al. (1994). *Plant Physiol.* **106**:447-458.

Opperman et al. (1993) *Science* **263**:221-223.

Overvoorde & Grimes (1994) *J. Biol. Chem.* **269**:1515415161.

Overvoorde et al. (1996) *Plant Cell* **8**:271280.

Pearson & Lipman (1988) *Proc. Natl. Acad. sci. USA* **85**: 2444

Pearson et al. (1994) *Methods in Molecular Biology* **24**, 307-31.

Pouwels et al. (1987) *Cloning Vectors: A Laboratory Manual*, 1985, supp.

Riesmeier et al. (1992) *EMBO J.* **11**:47054713.

Ripp et al. (1988) *Plant Physiol.* **88**:14351445.

Roshal et al. (1987). *EMBO J.* **6**:1155.

Saalbach et al. (1996) *Plant Physiol.* **112**:975-985.

Sambrook et al. (1989) In *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York.

Schaffner & Sheen (1991) *Plant Cell* **3**:997.

Schernthaner et al. (1988) *EMBO J.* 7:1249.

Schmidt (1994) *Science* **265**:739

Shewry et al. (1995) *The Plant Cell* 7:945-956.

Siebertz et al. (1989) *Plant Cell* **1**:961.

Smith & Waterman (1981) *Adv. Appl. Math.* **2**: 482

Southern (1975). *J. Mol. Biol.* **98**:503.

Southern et al. (1982) *J. Mol. Appl. Genet.* **1**:327341.

Stockhause et al. (1997) *The Plant Cell* **9**:479-489.

Terada & Shimamoto (1990) *Mol. Gen. Genet.* **220**:389.

Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York.

Weissbach & Weissbach (1989) Methods for Plant Molecular Biology, Academic Press.

SEQUENCE LISTING

(1) GENERAL INFORMATION
(i) APPLICANT: Grimes
(ii) TITLE OF INVENTION: Sucrose binding proteins
(iii) NUMBER OF SEQUENCES: 15
(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Klarquist Sparkman Campbell Leigh & Whinston, LLP
(B) STREET: One World Trade Center
121 S.W. Salmon Street
Suite 1600
(C) CITY: Portland
(D) STATE: Oregon
(E) COUNTRY: United States of America
(F) ZIP: 97204-2988
(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Disk, 3-1/2 inch
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: Windows NT
(D) SOFTWARE: Word 97 & ASCII
(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:
(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 60/047,568
(B) FILING DATE: May 22, 1997
(C) CLASSIFICATION:
(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: David J. Earp
(B) REGISTRATION NUMBER: 41,401
(C) REFERENCE/DOCKET NUMBER: 4630-50206/DJE
(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (503) 226-7391
(B) TELEFAX: (503) 228-9446

(2) INFORMATION FOR SEQ ID NO: 1:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 524
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Met Gly Met Arg Thr Lys Leu Ser Leu Ala Ile Phe Phe Phe Phe
 5                  10                  15

Leu Leu Ala Leu Phe Ser Asn Leu Ala Phe Gly Lys Cys Lys Glu
            20                  25                  30

Thr Glu Val Glu Glu Glu Asp Pro Glu Leu Val Thr Cys Lys His
        35                  40                  45

Gln Cys Gln Gln Gln Gln Gln Tyr Thr Glu Gly Asp Lys Arg Val
    50                  55                  60

Cys Leu Gln Ser Cys Asp Arg Tyr His Arg Met Lys Gln Glu Arg
65                  70                  75

Glu Lys Gln Ile Gln Glu Glu Thr Arg Glu Lys Lys Glu Glu Glu
            80                  85                  90
```

```
Ser Arg Glu Arg Glu Glu Glu Gln Gln Glu Gln His Glu Glu Gln
                 95                 100                 105

Asp Glu Asn Pro Tyr Ile Phe Glu Glu Asp Lys Asp Phe Glu Thr
            110                 115                 120

Arg Val Glu Thr Glu Gly Gly Arg Ile Arg Val Leu Lys Lys Phe
            125                 130                 135

Thr Glu Lys Ser Lys Leu Leu Gln Gly Ile Glu Asn Phe Arg Leu
            140                 145                 150

Ala Ile Leu Glu Ala Arg Ala His Thr Phe Val Ser Pro Arg His
            155                 160                 165

Phe Asp Ser Glu Val Val Phe Phe Asn Ile Lys Gly Arg Ala Val
            170                 175                 180

Leu Gly Leu Val Ser Glu Ser Glu Thr Glu Lys Ile Thr Leu Glu
            185                 190                 195

Pro Gly Asp Met Ile His Ile Pro Ala Gly Thr Pro Leu Tyr Ile
            200                 205                 210

Val Asn Arg Asp Glu Asn Asp Lys Leu Phe Leu Ala Met Leu His
            215                 220                 225

Ile Pro Val Ser Val Ser Thr Pro Gly Lys Phe Glu Glu Phe Phe
            230                 235                 240

Ala Pro Gly Gly Arg Asp Pro Glu Ser Val Leu Ser Ala Phe Ser
            245                 250                 255

Trp Asn Val Leu Gln Ala Ala Leu Gln Thr Pro Lys Gly Lys Leu
            260                 265                 270

Glu Asn Val Phe Asp Gln Gln Asn Glu Gly Ser Ile Phe Arg Ile
            275                 280                 285

Ser Arg Glu Gln Val Arg Ala Leu Ala Pro Thr Lys Lys Ser Ser
            290                 295                 300

Trp Trp Pro Phe Gly Gly Glu Ser Lys Pro Gln Phe Asn Ile Phe
            305                 310                 315

Ser Lys Arg Pro Thr Ile Ser Asn Gly Tyr Gly Arg Leu Thr Glu
            320                 325                 330

Val Gly Pro Asp Asp Asp Glu Lys Ser Trp Leu Gln Arg Leu Asn
            335                 340                 345

Leu Met Leu Thr Phe Thr Asn Ile Thr Gln Arg Ser Met Ser Thr
            350                 355                 360

Ile His Tyr Asn Ser His Ala Thr Lys Ile Ala Leu Val Ile Asp
            365                 370                 375

Gly Arg Gly His Leu Gln Ile Ser Cys Pro His Met Ser Ser Arg
            380                 385                 390

Ser Ser His Ser Lys His Asp Lys Ser Ser Pro Ser Tyr His Arg
            395                 400                 405

Ile Ser Ser Asp Leu Lys Pro Gly Met Val Phe Val Val Pro Pro
            410                 415                 420

Gly His Pro Phe Val Thr Ile Ala Ser Asn Lys Glu Asn Leu Leu
            425                 430                 435

Met Ile Cys Phe Glu Val Asn Ala Arg Asp Asn Lys Lys Phe Thr
            440                 445                 450
```

```
Phe Ala Gly Lys Asp Asn Ile Val Ser Ser Leu Asp Asn Val Ala
                455                 460                 465

Lys Glu Leu Ala Phe Asn Tyr Pro Ser Glu Met Val Asn Gly Val
                470                 475                 480

Phe Leu Leu Gln Arg Phe Leu Glu Arg Lys Leu Ile Gly Arg Leu
                485                 490                 495

Tyr His Leu Pro His Lys Asp Arg Lys Glu Ser Phe Phe Phe Pro
                500                 505                 510

Phe Glu Leu Pro Arg Glu Glu Arg Gly Arg Arg Ala Asp Ala
                515                 520
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 444

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Gly Met Arg Thr Lys Leu Ser Leu Ala Ile Phe Phe Phe Phe
                  5                  10                  15

Leu Leu Ala Leu Phe Ser Asn Leu Ala Phe Gly Lys Cys Lys Glu
                 20                  25                  30

Thr Glu Val Glu Glu Glu Asp Pro Glu Leu Val Thr Cys Lys His
                 35                  40                  45

Gln Cys Gln Gln Gln Gln Gln Tyr Thr Glu Gly Asp Lys Arg Val
                 50                  55                  60

Cys Leu Gln Ser Cys Asp Arg Tyr His Arg Met Lys Gln Glu Arg
                 65                  70                  75

Glu Lys Gln Ile Gln Glu Glu Thr Arg Glu Lys Lys Glu Glu Glu
                 80                  85                  90

Ser Arg Glu Arg Glu Glu Glu Gln Gln Glu Gln His Glu Glu Gln
                 95                 100                 105

Asp Glu Asn Pro Tyr Ile Phe Glu Glu Asp Lys Asp Phe Glu Thr
                110                 115                 120

Arg Val Glu Thr Glu Gly Gly Arg Ile Arg Val Leu Lys Lys Phe
                125                 130                 135

Thr Glu Lys Ser Lys Leu Leu Gln Gly Ile Glu Asn Phe Arg Leu
                140                 145                 150

Ala Ile Leu Glu Ala Arg Ala His Thr Phe Val Ser Pro Arg His
                155                 160                 165

Phe Asp Ser Glu Val Val Phe Phe Asn Ile Lys Gly Arg Ala Val
                170                 175                 180

Leu Gly Leu Val Ser Glu Ser Glu Thr Glu Lys Ile Thr Leu Glu
                185                 190                 195

Pro Gly Asp Met Ile His Ile Pro Ala Gly Thr Pro Leu Tyr Ile
                200                 205                 210

Val Asn Arg Asp Glu Asn Asp Lys Leu Phe Leu Ala Met Leu His
                215                 220                 225

Ile Pro Val Ser Val Ser Thr Pro Gly Lys Phe Glu Glu Phe Phe
                230                 235                 240

Ala Pro Gly Gly Arg Asp Pro Glu Ser Val Leu Ser Ala Phe Ser
                245                 250                 255
```

```
Trp Asn Val Leu Gln Ala Ala Leu Gln Thr Pro Lys Gly Lys Leu
                260                 265                 270

Glu Asn Val Phe Asp Gln Gln Asn Glu Gly Ser Ile Phe Arg Ile
            275                 280                 285

Ser Arg Glu Gln Val Arg Ala Leu Ala Pro Thr Lys Lys Ser Ser
        290                 295                 300

Trp Trp Pro Phe Gly Gly Glu Ser Lys Pro Gln Phe Asn Ile Phe
    305                 310                 315

Ser Lys Arg Pro Thr Ile Ser Asn Gly Tyr Gly Arg Leu Thr Glu
                320                 325                 330

Val Gly Pro Asp Asp Asp Glu Lys Ser Trp Leu Gln Arg Leu Asn
            335                 340                 345

Leu Met Leu Thr Phe Thr Asn Ile Thr Gln Arg Ser Met Ser Thr
        350                 355                 360

Ile His Tyr Asn Ser His Ala Thr Lys Ile Ala Leu Val Ile Asp
    365                 370                 375

Gly Arg Gly His Leu Gln Ile Ser Cys Pro His Met Ser Ser Arg
                380                 385                 390

Ser Ser His Ser Lys His Asp Lys Ser Ser Pro Ser Tyr His Arg
            395                 400                 405

Ile Ser Ser Asp Leu Lys Pro Gly Met Val Phe Val Val Pro Pro
        410                 415                 420

Gly His Pro Phe Val Thr Ile Ala Ser Asn Lys Glu Asn Leu Leu
    425                 430                 435

Met Ile Cys Phe Glu Val Asn Ala Arg
                440
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 489

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Ala Thr Arg Ala Lys Leu Ser Leu Ala Ile Phe Leu Phe Phe
 5                  10                  15

Leu Leu Ala Leu Ile Ser Asn Leu Ala Leu Gly Lys Leu Lys Glu
            20                  25                  30

Thr Glu Val Glu Glu Asp Pro Glu Leu Val Thr Cys Lys His Gln
        35                  40                  45

Cys Gln Gln Gln Arg Gln Tyr Thr Glu Ser Asp Lys Arg Thr Cys
    50                  55                  60

Leu Gln Gln Cys Asp Ser Met Lys Gln Glu Arg Glu Lys Gln Val
65                  70                  75

Glu Glu Glu Thr Arg Glu Lys Glu Glu Glu His Gln Glu Gln His
            80                  85                  90

Glu Glu Glu Glu Asp Glu Asn Pro Tyr Val Phe Glu Glu Asp Lys
        95                  100                 105

Asp Phe Ser Thr Arg Val Glu Thr Glu Gly Gly Ser Ile Arg Val
    110                 115                 120

Leu Lys Lys Phe Thr Glu Lys Ser Lys Leu Leu Gln Gly Ile Glu
```

```
                125                 130                 135

Asn Phe Arg Leu Ala Ile Leu Glu Ala Arg Ala His Thr Phe Val
            140                 145                 150

Ser Pro Arg His Phe Asp Ser Glu Val Val Leu Phe Asn Ile Lys
            155                 160                 165

Gly Arg Ala Val Leu Gly Leu Val Arg Glu Ser Glu Thr Glu Lys
            170                 175                 180

Ile Thr Leu Glu Pro Gly Asp Met Ile His Ile Pro Ala Gly Thr
            185                 190                 195

Pro Leu Tyr Ile Val Asn Arg Asp Glu Asn Glu Lys Leu Leu Leu
            200                 205                 210

Ala Met Leu His Ile Pro Val Ser Thr Pro Gly Lys Phe Glu Glu
            215                 220                 225

Phe Phe Gly Pro Gly Gly Arg Asp Pro Glu Ser Val Leu Ser Ala
            230                 235                 240

Phe Ser Trp Asn Val Leu Gln Ala Ala Leu Gln Thr Pro Lys Gly
            245                 250                 255

Lys Leu Glu Arg Leu Phe Asn Gln Gln Asn Glu Gly Ser Ile Phe
            260                 265                 270

Lys Ile Ser Arg Glu Arg Val Arg Ala Leu Ala Pro Thr Lys Lys
            275                 280                 285

Ser Ser Trp Trp Pro Phe Gly Gly Glu Ser Lys Ala Gln Phe Asn
            290                 295                 300

Ile Phe Ser Lys Arg Pro Thr Phe Ser Asn Gly Tyr Gly Arg Leu
            305                 310                 315

Thr Glu Val Gly Pro Asp Asp Glu Lys Ser Trp Leu Gln Arg Leu
            320                 325                 330

Asn Leu Met Leu Thr Phe Thr Asn Ile Thr Gln Arg Ser Met Ser
            335                 340                 345

Thr Ile His Tyr Asn Ser His Ala Thr Lys Ile Ala Leu Val Met
            350                 355                 360

Asp Gly Arg Gly His Leu Gln Ile Ser Cys Pro His Met Ser Ser
            365                 370                 375

Arg Ser Asp Ser Lys His Asp Lys Ser Ser Pro Ser Tyr His Arg
            380                 385                 390

Ile Ser Ala Asp Leu Lys Pro Gly Met Val Phe Val Val Pro Pro
            395                 400                 405

Gly His Pro Phe Val Thr Ile Ala Ser Asn Lys Glu Asn Leu Leu
            410                 415                 420

Ile Ile Cys Phe Glu Val Asn Val Arg Asp Asn Lys Lys Phe Thr
            425                 430                 435

Phe Ala Gly Lys Asp Asn Ile Val Ser Ser Leu Asp Asn Val Ala
            440                 445                 450

Lys Glu Leu Ala Phe Asn Tyr Pro Ser Glu Met Val Asn Gly Val
            455                 460                 465

Ser Glu Arg Lys Glu Ser Leu Phe Phe Pro Phe Glu Leu Pro Ser
            470                 475                 480

Glu Glu Arg Gly Arg Arg Ala Val Ala
            485
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 409

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Ala Thr Arg Ala Lys Leu Ser Leu Ala Ile Phe Leu Phe Phe
                5                  10                  15

Leu Leu Ala Leu Ile Ser Asn Leu Ala Leu Gly Lys Leu Lys Glu
            20                  25                  30

Thr Glu Val Glu Glu Asp Pro Glu Leu Val Thr Cys Lys His Gln
        35                  40                  45

Cys Gln Gln Gln Arg Gln Tyr Thr Glu Ser Asp Lys Arg Thr Cys
    50                  55                  60

Leu Gln Gln Cys Asp Ser Met Lys Gln Glu Arg Glu Lys Gln Val
65                  70                  75

Glu Glu Glu Thr Arg Glu Lys Glu Glu Glu His Gln Glu Gln His
                80                  85                  90

Glu Glu Glu Glu Asp Glu Asn Pro Tyr Val Phe Glu Glu Asp Lys
            95                 100                 105

Asp Phe Ser Thr Arg Val Glu Thr Glu Gly Gly Ser Ile Arg Val
        110                 115                 120

Leu Lys Lys Phe Thr Glu Lys Ser Lys Leu Leu Gln Gly Ile Glu
    125                 130                 135

Asn Phe Arg Leu Ala Ile Leu Glu Ala Arg Ala His Thr Phe Val
140                 145                 150

Ser Pro Arg His Phe Asp Ser Glu Val Val Leu Phe Asn Ile Lys
                155                 160                 165

Gly Arg Ala Val Leu Gly Leu Val Arg Glu Ser Glu Thr Glu Lys
            170                 175                 180

Ile Thr Leu Glu Pro Gly Asp Met Ile His Ile Pro Ala Gly Thr
        185                 190                 195

Pro Leu Tyr Ile Val Asn Arg Asp Glu Asn Glu Lys Leu Leu Leu
    200                 205                 210

Ala Met Leu His Ile Pro Val Ser Thr Pro Gly Lys Phe Glu Glu
215                 220                 225

Phe Phe Gly Pro Gly Gly Arg Asp Pro Glu Ser Val Leu Ser Ala
                230                 235                 240

Phe Ser Trp Asn Val Leu Gln Ala Ala Leu Gln Thr Pro Lys Gly
            245                 250                 255

Lys Leu Glu Arg Leu Phe Asn Gln Gln Asn Glu Gly Ser Ile Phe
        260                 265                 270

Lys Ile Ser Arg Glu Arg Val Arg Ala Leu Ala Pro Thr Lys Lys
    275                 280                 285

Ser Ser Trp Trp Pro Phe Gly Gly Glu Ser Lys Ala Gln Phe Asn
290                 295                 300

Ile Phe Ser Lys Arg Pro Thr Phe Ser Asn Gly Tyr Gly Arg Leu
                305                 310                 315

Thr Glu Val Gly Pro Asp Asp Glu Lys Ser Trp Leu Gln Arg Leu
```

```
                320                 325                 330

Asn Leu Met Leu Thr Phe Thr Asn Ile Thr Gln Arg Ser Met Ser
                335                 340                 345

Thr Ile His Tyr Asn Ser His Ala Thr Lys Ile Ala Leu Val Met
                350                 355                 360

Asp Gly Arg Gly His Leu Gln Ile Ser Cys Pro His Met Ser Ser
                365                 370                 375

Arg Ser Asp Ser Lys His Asp Lys Ser Ser Pro Ser Tyr His Arg
                380                 385                 390

Ile Ser Ala Asp Leu Lys Pro Gly Met Val Phe Val Val Pro Pro
                395                 400                 405

Gly His Pro Phe
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1924

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
TGTAAAACGA CGGCCAGTGA ATTGTAATAC GACTCACTAT AGGGCGAATT          50
GGGTACCGGG CCCCCCCTCG AGGTCGACGG TATCGATAAG CTTGATTTTG         100
TTCCTCACTG ACCTCACC  ATG GCG ACC AGA GCC AAG CTT TCT TTA       145
                     Met Ala Thr Arg Ala Lys Leu Ser Leu
                                      5
GCT ATC TTC CTT TTC TTT CTT TTA GCC TTG ATT TCA AAC CTA GCC   190
Ala Ile Phe Leu Phe Phe Leu Leu Ala Leu Ile Ser Asn Leu Ala
 10                  15                  20

TTG GGC AAA CTT AAA GAA ACC GAG GTC GAA GAA GAT CCC GAG CTC   235
Leu Gly Lys Leu Lys Glu Thr Glu Val Glu Glu Asp Pro Glu Leu
 25                  30                  35

GTA ACA TGC AAA CAC CAG TGC CAA CAG CAA CGG CAA TAC ACT GAG   280
Val Thr Cys Lys His Gln Cys Gln Gln Gln Arg Gln Tyr Thr Glu
 40                  45                  50

AGT GAC AAG CGA ACA TGC TTG CAA CAA TGT GAC AGT ATG AAG CAA   325
Ser Asp Lys Arg Thr Cys Leu Gln Gln Cys Asp Ser Met Lys Gln
 55                  60                  65

GAG CGA GAG AAA CAA GTC GAA GAG GAA ACT CGC GAG AAG GAA GAA   370
Glu Arg Glu Lys Gln Val Glu Glu Glu Thr Arg Glu Lys Glu Glu
 70                  75                  80

GAA CAT CAA GAG CAG CAT GAG GAG GAG GAA GAC GAA AAT CCC TAC   415
Glu His Gln Glu Gln His Glu Glu Glu Glu Asp Glu Asn Pro Tyr
 85                  90                  95

GTT TTT GAA GAA GAT AAG GAT TTT TCG ACC AGA GTC GAA ACA GAA   460
Val Phe Glu Glu Asp Lys Asp Phe Ser Thr Arg Val Glu Thr Glu
100                 105                 110

GGT GGC AGC ATT CGG GTT CTC AAG AAG TTC ACT GAG AAA TCC AAG   505
Gly Gly Ser Ile Arg Val Leu Lys Lys Phe Thr Glu Lys Ser Lys
115                 120                 125

CTT CTT CAA GGC ATT GAG AAT TTC CGT TTG GCC ATC TTA GAA GCT   550
Leu Leu Gln Gly Ile Glu Asn Phe Arg Leu Ala Ile Leu Glu Ala
130                 135                 140
```

```
AGA GCA CAC ACG TTC GTG TCC CCA CGC CAC TTT GAT TCC GAG GTT  595
Arg Ala His Thr Phe Val Ser Pro Arg His Phe Asp Ser Glu Val
145                 150                 155

GTC TTG TTC AAC ATT AAG GGG AGA GCC GTA CTT GGG TTG GTG AGG  640
Val Leu Phe Asn Ile Lys Gly Arg Ala Val Leu Gly Leu Val Arg
160                 165                 170

GAA AGT GAA ACA GAA AAA ATC ACC CTA GAA CCT GGA GAC ATG ATA  685
Glu Ser Glu Thr Glu Lys Ile Thr Leu Glu Pro Gly Asp Met Ile
175                 180                 185

CAC ATA CCA GCA GGC ACA CCA CTG TAC ATC GTT AAC AGA GAT GAG  730
His Ile Pro Ala Gly Thr Pro Leu Tyr Ile Val Asn Arg Asp Glu
190                 195                 200

AAT GAG AAG CTC CTC CTT GCC ATG CTC CAT ATA CCT GTC TCT ACT  775
Asn Glu Lys Leu Leu Leu Ala Met Leu His Ile Pro Val Ser Thr
205                 210                 215

CCT GGA AAA TTT GAG GAA TTT TTC GGG CCT GGA GGA CGA GAC CCA  820
Pro Gly Lys Phe Glu Glu Phe Phe Gly Pro Gly Gly Arg Asp Pro
220                 225                 230

GAA TCG GTC CTC TCA GCA TTC AGC TGG AAT GTG CTG CAA GCT GCG  865
Glu Ser Val Leu Ser Ala Phe Ser Trp Asn Val Leu Gln Ala Ala
235                 240                 245

CTC CAA ACC CCA AAA GGA AAG TTA GAA AGG CTT TTT AAT CAA CAG  910
Leu Gln Thr Pro Lys Gly Lys Leu Glu Arg Leu Phe Asn Gln Gln
250                 255                 260

AAC GAG GGA AGT ATT TTC AAA ATA AGC AGA GAA CGG GTG CGT GCG  955
Asn Glu Gly Ser Ile Phe Lys Ile Ser Arg Glu Arg Val Arg Ala
265                 270                 275

TTG GCC CCC ACC AAG AAA AGC TCT TGG TGG CCA TTC GGC GGC GAA 1000
Leu Ala Pro Thr Lys Lys Ser Ser Trp Trp Pro Phe Gly Gly Glu
280                 285                 290

TCC AAG GCT CAA TTC AAT ATT TTC AGC AAG CGT CCC ACT TTC TCC 1045
Ser Lys Ala Gln Phe Asn Ile Phe Ser Lys Arg Pro Thr Phe Ser
295                 300                 305

AAC GGA TAT GGC CGT TTA ACT GAA GTT GGT CCT GAT GAT GAA AAG 1090
Asn Gly Tyr Gly Arg Leu Thr Glu Val Gly Pro Asp Asp Glu Lys
310                 315                 320

AGT TGG CTT CAA AGA CTC AAC CTC ATG CTT ACC TTT ACC AAC ATC 1165
Ser Trp Leu Gln Arg Leu Asn Leu Met Leu Thr Phe Thr Asn Ile
325                 330                 335

ACC CAG AGA TCT ATG AGT ACT ATT CAC TAC AAC TCA CAT GCA ACG 1180
Thr Gln Arg Ser Met Ser Thr Ile His Tyr Asn Ser His Ala Thr
340                 345                 350

AAG ATA GCA CTG GTG ATG GAT GGT AGA GGG CAT CTT CAA ATA TCA 1225
Lys Ile Ala Leu Val Met Asp Gly Arg Gly His Leu Gln Ile Ser
355                 360                 365

TGT CCA CAC ATG TCA TCA AGG TCA GAC TCA AAG CAT GAT AAG AGT 1270
Cys Pro His Met Ser Ser Arg Ser Asp Ser Lys His Asp Lys Ser
370                 375                 380
```

```
AGC CCC TCA TAC CAT AGA ATC AGT GCG GAC TTG AAG CCT GGA ATG 1315
Ser Pro Ser Tyr His Arg Ile Ser Ala Asp Leu Lys Pro Gly Met
385                 390                 395

GTG TTT GTT GTC CCT CCT GGT CAT CCC TTC GTC ACT ATA GCT TCC 1360
Val Phe Val Val Pro Pro Gly His Pro Phe Val Thr Ile Ala Ser
400                 405                 410

AAT AAA GAG AAT CTC CTC ATA ATT TGC TTC GAG GTT AAC GTT CGA 1405
Asn Lys Glu Asn Leu Leu Ile Ile Cys Phe Glu Val Asn Val Arg
415                 420                 425

GAC AAC AAG AAG TTT ACG TTT GCA GGG AAG GAC AAC ATT GTG AGC 1450
Asp Asn Lys Lys Phe Thr Phe Ala Gly Lys Asp Asn Ile Val Ser
430                 435                 440

TCT CTG GAC AAC GTA GCT AAG GAG CTG GCC TTT AAC TAT CCT TCT 1495
Ser Leu Asp Asn Val Ala Lys Glu Leu Ala Phe Asn Tyr Pro Ser
445                 450                 455

GAG ATG GTG AAC GGA GTC TCC GAA AGA AAG GAG AGT CTC TTT TTC 1540
Glu Met Val Asn Gly Val Ser Glu Arg Lys Glu Ser Leu Phe Phe
460                 465                 470

CCC TTC GAG TTG CCG AGC GAG GAG CGT GGT CGT CGC GCT GTT GCG 1585
Pro Phe Glu Leu Pro Ser Glu Glu Arg Gly Arg Arg Ala Val Ala
475                 480                 485

TGA GAAGCAGTGT GGAGGTGGCT GATAACGGGG AATGTATTTA GCTTTGAGAG 1638
TCTTTAAATT TTCTGTATTT GTTGTAATGT TAGTAGTTCC TTAAATTGGC      1688
CAGATGGAGT TTATGTGTTT GTAAATGCAG GGATGCTAAC GGAATAAAAT      1738
GGCCACTTGT ATTGCTAAAG AAAAAAACCA GCCCGGGCCG TCGACCACGC      1788
GTGCCCTATA GTGAGTCGTA TTACAATCGA ATTCCTGCAG CCCGGGGGAT      1838
CCACTAGTTC TAGAGCGGCC GCCACCGCGG TGGAGCTCCA GCTTTTGTTC      1888
CCTTTAGTGA GGGTTAATTT CGAGCTTGGC GTAATC                     1924
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3718

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

i SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
TTGTAAACGA CGGCCAGTGA ATTGTAATAC GACTCACTAT AGGGCGAATT      50
GGGTACCGGG CCCCCCCTCG AGGTCGACGG TATCGATAAG CTTGATTGTA     100
ATACGACTCA CTATAGGGCA CGCGTGGTCG ACGGCCCGGG CTGGTCTGAG     150
AAACTCATTA GGCACTGGAA AATTCTCAAA GGAAATAATG TGAGTCAGCC     200
AATTCAAACC CACCATATCT TTATTAATTT CACTTTTTTC TTTATTTTAT     250
AATTTTTAGT CTCACAGTCA CACATTTTAA CAGGTTATGA TAACAAGGGG     300
CAAAGATAAG GGTGAGACCG GGATTATAAA GCGTGTCATT CGCTCTCAAA     350
ATCGTGTCAT TGTAGAGAGT AAAAACCTGG TGAGAGATAT TATCATCACA     400
ATTTGGTCCT TCTGTTTTTC TAATGCCCTA TCTTCCTTAG ATTATGTTTT     450
CAATTCCACT GTCAATGTGT CTTGCATCAG AATATTAATC AATTGTGACA     500
TTGAGCATGT GATTGTGTAA ATTTTCCTGA TAGGTTTCTC ACTCCAATGC     550
CTTTTGTCAT CCTCTTTATA GGTAAAGAAG CATATAAAGC AAGGGCAAGG     600
TCATGAAGGG GGCATCTTTA CAGTGGAAGC CCCACTGCAT GCCTCCAATG     650
TGCAAGTTCT TGACCCAGTG ACAGGGTATG TCATTGTTCA GATATTGAAC     700
TGGTGATTGC ATCTCCAAAC GGGATAACAT CATTAACATG TATGAAAGTA     750
AGAGTTACCA ACTTTTACTT GTGCAGCAAG CCTTGCAAGG TTGGAGTTAA     800
ATATCTTGAA GATGGTACTA AAGTCAGAGT GTCCAGAGGA ATAGGAACCT     850
CAGGGTCCAT AGTCCCTCGT CCTGAGATTT TAAAGATAAG AACTACCCCA     900
AGACCTGCAG TCCGTAAGTA TCTAACAAGC TTAATTATGC TTTTTCATGT     950
ATGAGTTGTT GACAAAACAT GGCCAGAGCC AATAGAGAAT CGAGAAAAAG    1000
TGAGACGGAA AATGAACTTG AATTATGAGA AAGGTGTGTG AAACAAACAA    1050
```

GCCAATAATG TGGCTTATAT AATATATAAT ATATAGATAT AGACCAGAGT 1100
GAGTAACGAA TCACTAACTA ATTACATGTG TATATCTACC TAATTAGATG 1150
ACTCATCAAA CAAAGCGAAC TATTGTGATA GAGACTTTAT TTTTCGCAAT 1200
TAATTCAAAG ATGTACTGCT TATCTTCTTT GCTACATGTC TGTTGACATG 1250
CATTGTTATC CATAACCTTG TTATTATACT TGGTGTTGAG AAAGGAGAGT 1300
CTCCTTGCAC TTTAGAGACA TTCTTTAAAC TGACTTGACC TTATTGAAAA 1350
TTCGAGATAG CAACTTAGCA CCACACCTTA AAAAGAAAGA TTTTTTAGAG 1400
GGTAGATTAA TTGTTGAATA ATGTTAATCA TCAAAGGTTT AAGATTTATT 1450
AAGTGCTTTC CATTGTCTTA AAAATCTTGC TTCTAGGACT AGGATGTGTA 1500
TTGTTACATG ATTTCCCCCC CTTGGTATCA ACTAAAGCAT GTTGGACTTG 1550
CGCTCCATAT GCAGAAACTC AAATTAAAAA CATCATTTGT AATGTATAGT 1600
AAGTGTATAT ATAACATTGT AAGTTGTCGA TCAAAGTTAT TTGGATTAAT 1650
GGATTTAAGT CTTCTATAAT ATTCCATTGA GAGCCAGAAG CCAGGTCCAA 1700
AGGAATAAGT AACTCGCATG AATTCATTCT CTTGCTTCTA TACAGCTATT 1750
TTTCCATCTT AGTGTTGCGG GAAACTACTT CAGTTCTCGC AGATGTGCAA 1800
AACTTGTAGG GATCCATGTA GTTCAGTGAA ACCCATGCTT TCTTAATTGA 1850
CAGAGATACA TTAAAACTTT TTACAGAATT GAGAAACCCA AGCCTTGTTA 1900
ATTCTCAAAG ATACATTTAA ACTTTTTTCA GAAACGTGCT GAGTATTTTA 1950
TCCTGTTTGT TATTCATTTT TGGCAGTTGG TCCTAAAAAT ACTCCTATGA 2000
ATCTTGTGCT AGAGAAGACT TGCAATGCTA AAACAGGACG GGGCATGCCT 2050
GAACTTTAAG GAGACGTTGC CTTGTTCCGA TTAGGTAATT GCTATCGTGA 2100
TGAACAAAAA TTTGGTGTGA ATTTATCCCC TTGCCCTTTG CCATGATTCA 2150
ATTAAAGACG TGTTTGGAAC CACATTCTAA CACCACTTTA TGATGGGTTA 2200
GACGCAAAAT CTAGATTGGG TAGTGTTTAC ACACAGTTAC AAACACATTC 2250
CTTGTTTAAT GTTATCATGC CTAGGAGTTG AATAACTTGT AACTTTACCA 2300
ATTAGACATT ACTACTAGCA TTCTTTTTCC TATTCAAGTT GATGTTATCT 2350
CCAGTTAGTG ATGGTCATTT CATTCCATAA ACTTCAATTG TTAAAATGAG 2400
TGAAAAGGGA AAAAGGAACC CGTTTGATTG TTATGGTTCT AGTGATTTTT 2450
ATTAATTGGG TTTGTCCATT AGTGTCGATT TGAGCTAAAT AGTTTCCCCC 2500
CCCCAAAAGA TCAGTCTTCT CACATGTCAT ATTCATGCGC TGGTACCCTT 2550
TTCATCCAGT TCCAACAAAC TTGCTGTACG AAGTCAGGTT GCATGAAAAT 2600
AGTCAAATTT TCTTTAAGGG GGATATTATA CGTAAATAAA TAACGTAACC 2650
CAAAAGTCTT ACTTGTTGGG TAACGTGGGT TTTGGTGTTT GATGGACCTA 2700
GAACACTGTT TGTTGCTCTT ATATGCTTAC AAAGTAAAAA TGGTTATCAC 2750
ATTTGGGGAA AAAATGTAGG CCCACTTATG ATATTTCGAC CTAAATGCAA 2800
AATGGTTTAT CAATTTTTTT ATACTTAGTA TGATAAAACT CCTTTTTTTT 2850
TTCCACTGGC ATACTATTTC TCTAAGACTT TTTAATAGTT CCGATAATTC 2900
TTAGCTTAAA GAAATACGAC AAGGTTAGGA ATATTTTTTT ATTATGTGAC 2950
ATTATTTTTT AAATATTTTG CTTCATATGA ATTTATACAA TCATTATAAT 3000
TTGACCTTTT AAATGACTTT TAAAAATGAT CAGACCTAAA ATTTGAGTCT 3050
TCTGATTGAG ATGCAAACTT ATTTCTTTTT ATATTTTATA TTTTATACTC 3100
ATTTGTTTCT CTTTCTATTA TATTTCTTTT TTTTCTTCTC TTTATGCAAA 3150
AACGTATGAC GTTGATTGGT GTCTTTGGCA ATCTTTTTAT GACGCTCAAA 3200
AGTGAAAATA AATATTGTTC ACTTTCACCT CACGCTGGCC TTCCGCTGAT 3250
GGTGGTTGTA CGCACTTATT TGATTTTTTT TTCTTCCACA TTTAATGAGG 3300
TGAATCAGTT AGAGAAATAT TAAAAAAAAT AAATAAATAA AGGAAGACGA 3350
CTAATACAAT AAAGAATACG AAACTCACAA TGAATAGACC CAATTAGAAC 3400
CATTTATTTT CCTTACAAAT TAAAGAAAAC GTTTTTTTAA CAATATATCA 3450
CATTATCATC TATTATATTT TTATTTATAT TTTTTATAAC TTTCTCTATC 3500
TAGGTGTAGA TTGACATGAG TATACGCACG CACACCCAGC TCTACTTAGC 3550
AGCAATTACC CGTTTTACTT GCTACTTAAG AGACACGTAC ATTAACACTT 3600
GTCCTTGTGC ATGCAATTGC CACCACATTC CTCACTCCAC CCTTTTCTTT 3650
ATATATAAAC AAACACAATG GATCATCTCA AACCAAGAGT GAGTTTGTTG 3700
TTCCTCACTG ACCTCACC 3718

(2) INFORMATION FOR SEQ ID NO: 7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4476
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TTGTAAAACG ACGGCCAGTG AATTGTAATA CGACTCACTA TAGGGCGAAT 50

TGGGTACCGG GCCCCCCCTC GAGGTCGACG GTATCGATAA GCTTGATTAC 100
TATAGGGCAC GCGTGGTCGA CGGCCCGGGC TGGTACTTTT GACTCCCTAA 150
TTGACAACTA CTGCATTGTA TCGATATTAA TATGGAATTT GGAATCATGG 200
TCCATGCTTC ATGTATTGTG TACCTCATAT TCAACAGCTA GTGAACACAA 250
AATCTTACAT ACTTTTGTAT TTCTATCAGT TTATACCTTC CCAAATAAAT 300
GGCTTATATT GCATTGAGTT ACATATTATT GTTTAGTTGG ATTGTAATTT 350
ACGAGTAGTT TGTCACGACT GAAGAAATTA ATAAGGTATA AGACACGTCC 400
TGCTCCCGCG AAATTCATTT TCTGTTTATT CTCTGTCTCT GTCTCTATTC 450
AATTCAACCT TCCATTTGTT TTCGCCAGCA TCCAGATTTG TGCTTTCTCT 500
ATCATTTCAT TTAATTAATG TGATGTATGT ATGGCTGAAT AAAAGATGGA 550
TTCCTCTTTT TTGTGGGGTG GAAGCTTAAT CTATGGGGCT AGATAAAAAA 600
ATTCATCTGT TTGTTGCACA GAATAAAATA TAAATTAATA ATTAATTAAA 650
CTTCAAACAT GGACAGGGCA CCTCCAAGTT ATTTTAAAAC CGACCATGGC 700
CATTTTTGCT TTCTGTTGGT GTTCTTGGCT CAGCTTTTGT AATTTTAGAC 750
TGCAGAAACA TCCTGTATGG GTTGGAAAGC AGCTGAGAAA CTCATTAGGC 800
GCTGGAAAAT TCTAAGAGGG GATAATGTAT GTGAGTCAAT TCAAACCCAC 850
CATATGTTTG TCTCTGTGCT CTTTATTAAT TTCACTTTTT TATTTTATAA 900
TTTTAGTCTC ACAGTCACAG AGTCACTTAT GTATTCATCT AACAGGTTAT 950
GATAACAAGG GGCAAAGATA AGGGTGAGAC CGGGATTATA AAGCGTGTCA 1000
TTTGCTCTCA AAATCGTGTC ATTGTAGAGG GTAAAAATCT GGTGAGATAT 1050
TATAATCACT ATTTGGTCCT TCTGTTTTTC TAATGCCCTA TCTTCTGTAG 1100
CTTTTGTTTT CAATTCCACT GTCAGTGTGT CTTGCATCAG AATATTAATC 1150
GGTTGTCAGT GACATTGAGC ATTTAATTGT GTAAATTTTC CTGTTAGATT 1200
TCTCACTCCA ATGCCTTTTG CCGTCCTCTT TATAGGTAAA GAAGCATATC 1250
AAGCAAGGGC AAGGTCATAA AGGGGGAATC TTTACAGTGG AAGCCCCACT 1300
GCATGCCTCC AATGTGCAAG TTCTTGACCC AGTGACAGGG TATGTACATG 1350
TTAGATATTG AACTGGTGAT TGCTTCTCCA AATGGGATAA CATGTATGTA 1400
AGTAAGAGTA ACCTACTTTT ACTTGTGCAG CAAGCCTTGC AAGGTTAGAG 1450
TTAAAATATC TTGAAGATGG TACTAAAGTC AGAGTGTCCA GAGGAATAGG 1500
AGCATCAGGG TTCATAGTCC CTCGTCCCAA GATCTTAAAG ATAAGAACTA 1550
CCCCAAGACC TACAGTCCGT AAGTATCTAA CAAGCTTATG TTTTTTCCTT 1600
GTATGAGTTG TTGATAAAAC ATGGCCAGAG CCAATAGAGA ATTGAGAAAA 1650
GGTGAGAAAC AGAAAATGAA CTTGAATTAT GAGAAAGGTG TGGGAAACAA 1700
ACAAGCCAAT AATGTGGCTT ATATAATATA TAGATATAGA CTAGAGTGAG 1750
TAACGAATCA CTAACTAATT ACATGTGCAT ATCTACCTAA TTAGATGATT 1800
CGTCAAACGA AGCAAAGTAT TGTGATAGAT AGTTGATTTT TCTCAAATAA 1850
TTCTAAGATG TAATACTTAT ATTCTTTGCT ACATGTCTGT TGACATACAT 1900
TGTTATCCAT AACCTTGTTA TTATACTTGG TGTTAAAAAA GGAGAGTCTC 1950
CTTGCACTTT AGAGACATTC TTTAAACTGA CTTGACCTTA TTGAAATACA 2000
TAATTCTAGT TACCAACTTA GCACCACACC ATAAAAGGAA AGATTTTTAA 2050
ACGGTAGATT GATTGTTGAA TAATGTTAAT CATCAAAGGT TTAAGATTTA 2100
TTAAGTGCTT TCCATTGTCT TAAAATATTG CTTCTAGGAC TAGGATGTGT 2150
ATATTGGTTA CATGATTTCC CCGCCTTCGT ATCAACTTAA GCATGTTGGA 2200
CTTGCACCCA TATGCAGAAA CTCAAATAAA AAACTTCATT TGTAAGGTAT 2250
AATAAGTGTA TATATAACAT TGTAAGTTGT CAATCAGAGT AATTTGGATT 2300
GATGGATATT TAAGTCTTCT ATAATATTTC ATTTAGAGCC AGAAGCCAGG 2350
TTCAAAGGAA TAGGTAATTC ACATGAATTC ATTCTCTTGT TTCTATACAG 2400
TTATTATTTT TTCCATCTTA GTGTTGCAGG AAACTACCTC AGTTGTTGTA 2450
GATGTGCAAA ACTTGTATGG ATATATATAC TGTTCAGTGT TGGGAAACCC 2500
ATGCTTTCTT AATTCACAGA GATACATTTA AACTTTTTTT AGAAACTTGC 2550
TTAGTATCTT ATCCTGTTAT TCATTTTTGG CAGTTGGTCC TAAAGATACT 2600
CCTATGAATC TTGTGCTAGA GAAGACTTAC GATGCTAAAA CAGGACGGGG 2650
CATGCCTGAA CTTTAAGGAG ACGTTGCCCT GTTCCACTTC CAATTAGGTA 2700
ACTGCTATCG TGATGAACAA AAATTTGGTG TGAGTTTATC ACCTTGTCCT 2750
TTGCCATGAT TCAATTAAAA GCGTGTTTGG ACTTTGGAAC CTCATTCTAA 2800
CACCACCCTA TGATGGGTTA GACGCAAAAT CTAGACTGGG TAGTGTTTAA 2850
CGTGTATCTG TGTGAACACA GTTACAAACG CATTCCATGT TTAATGCTAC 2900
CATGCCTAGG AGTTGAATCA TTTGTAACTT TACCAATTTA GTCATTACTA 2950
CTAGCATTCT TTTCCCTATT CAAGTTGATG TTAGCTCCAG TTAGGGATGG 3000
TCATTTCACT CCATAAACTT TAATTGTTAG GTGAGTGGAA GAGGAACCCG 3050
TTTGATTGTT ATGGTTCTAG TTCTAGTGAT TTTTATTAAT TGGGTTCGAC 3100
CATATTAGTG TTTGATTTGA GCTATAGATA GTTTTTTCCC CAAAAGATCA 3150
GTTTTCTCAC ATGTCAGATT CATGGGTTGG TACTCTTTTC ATCCAGTTCC 3200
AACAAACTTG CTGTTCGAAC TACGAAGTCA GTCTTACTTA TTGGGTAACA 3250
TGTGGGTTTT GGTGTTTAAT GGATCTAGAA TACTGTTTGT AGCTAAACCT 3300

```
ATCTTATCAT ATAGGGCCTA AAAAGTAAAA TTGGTTATTA CATTTGGAAA   3350
AAAAGAAATA ATCTAGGCCC ACTGGCACAC TGAAAAACGT TTTCAATGAA   3400
TAATTTAATA GTTTTTTTTT TATAAAAAAA TTTTAATAAA AAATAATGGA   3450
GTTTTTAAAA ATATTACAAC AATCTGTTTC TCTAAGGTTT TTTAATAGTT   3500
CAGATAATTC ATAGCTTAGA GCAATACGAC ATGGTTAGGA AGCATAAAAA   3550
AAATATACGA CATGGTTAGG AATTTTTTTT TAGTATGTCT GACATAATTT   3600
TTTAAATGTT TTGGCTTCAT ATGAATTTAA CAGTGCGTCA TATGAACTTA   3650
CACACTCATT ATATTTTTTA ACCTTTTAAA TGATTTTTAA AAAATATGAC   3650
AGATGCAATC TTATTCTCAC TTTTTATACT TTCACTACTG CTTCATATGA   3700
CCTAAAGTCA GAGAAATATT TTAAAAAGAT AAATACGATA AAGAATACGA   3750
TGAGAAAGAA ACCTCACACA ATGAATAGAC CAAATTAGAC CTATTTATTT   3800
TCCTTAGAAA TAAAGAAAAT AATTATTTTT TATTTTTTCA CATTACATTT   3850
ATATTTTTCT ATCACTTTCT CTATTTAGGT ATTGATTGAC ATATGAGTGT   3900
ACATGAACTT TTTTTAAAAA AAAAGCGTAA ATATTAATTA TATTCATGCA   3950
TTTGTTTTCT GTCTTTCATT TTCTATTTAA TCTTACGTTA TCAATAATCT   4000
ATTATTAAAT TTTATAGTTG ATGATGAATA TATAAGAGAT ATAAATAAAA   4050
AAATAATTAA TTTTATAATA AAAATTAAAA AATAATTAAT TATTTTGAGA   4100
TAAATTTTTT TTAAGAGAAC AATTATAAAC GGAGAGTATT ATATTTAGTT   4150
TTATGTGTAC CGGGTACGTG TCTACTAACA TGGTGTCTCT CCATCATTTT   4200
CGTAGGAAAA AACATTATAG GAGTATGAAA AAAGCAAAAG TTTTGTCTGT   4250
TTATGGTTTT GTATATACCC AGCTCTACTT GGCAGCAATT ACCCGTCTTG   4300
CTTGCTACTT ACGAGACACG TACATTAACA CTTGTCCTAG CTAGTGCATG   4350
CAATTGCCAC CCCATTCCTC ACTCCTCCCT TTTCCTTCTC TTTATATTTA   4400
TATATATAAA TAAACAAACA CAATGCATCA TCTCAAAGAA ATTAAGAGAG   4450
TTTTTTTGTT CCTCACTGAC CAAGCC                            4476
```

(2) INFORMATION FOR SEQ ID NO: 8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
TGTAAAACGA CGGCCAGTGA ATT 23

(2) INFORMATION FOR SEQ ID NO: 9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
GATTACGCCA AGCTCGAAAT TAA 23

(2) INFORMATION FOR SEQ ID NO: 10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO:10:
ATGGCGACCA GAGCCAAGCT TTCTTTA 27

(2) INFORMATION FOR SEQ ID NO: 11:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CGCAACAGCG CGACGACCAC GCTCGCT 27

(2) INFORMATION FOR SEQ ID NO: 12:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
ATGGCGACCA GAGCCAAGCT TTCTTTA 27

(2) INFORMATION FOR SEQ ID NO: 13:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
GAAGGGATGA CCAGGAGGGA CAACAAA 27

(2) INFORMATION FOR SEQ ID NO: 14:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
TTGTAAACGA CGGCCAGTGA ATT 23

(2) INFORMATION FOR SEQ ID NO: 15:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:
GGTGAGGTCA GTGAGGAACA ACA 23

CLAIMS

1. A modified plant sucrose binding protein wherein the modified sucrose binding protein has a modified amino acid sequence compared to a corresponding wild-type sucrose binding protein, and wherein expression of the modified sucrose binding protein in a yeast assay system confers enhanced sucrose compared to the corresponding wild-type sucrose binding protein.

2. A modified plant sucrose binding protein according to claim 1 wherein the modified sucrose binding protein enhances sucrose uptake in the yeast assay system by at least 10% compared to the wild-type sucrose binding protein.

3. A modified plant sucrose binding protein according to claim 1 wherein the modified sucrose binding protein enhances sucrose uptake in the yeast assay system by at least 25% compared to the wild-type sucrose binding protein.

4. A modified plant sucrose binding protein according to claim 1 wherein the modified amino acid sequence comprises a C-terminal truncation compared to the wild-type sucrose binding protein.

5. A modified plant sucrose binding protein according to claim 4 wherein the C-terminal truncation results in removal of between 10 and 100 amino acids.

6. A modified plant sucrose binding protein, wherein the corresponding wild-type sucrose binding protein is selected from the group consisting of SBP1 and SBP2.

7. A modified plant sucrose binding protein according to claim 6 wherein the protein has an amino acid sequence selected from the group consisting of Seq. I.D. Nos. 2 and 4.

8. A nucleic acid molecule encoding a modified plant sucrose binding protein according to claim 1.

9. A vector comprising a nucleic acid molecule according to claim 8.

10. A transgenic plant expressing a modified plant sucrose binding protein according to claim 1.

11. A nucleic acid molecule encoding a modified sucrose binding protein according to claim 6.

12. A transgenic plant expressing a modified plant sucrose binding protein according to claim 6.

13. An isolated nucleic acid molecule encoding a plant sucrose binding protein, wherein the protein comprises an amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth in Seq. I.D. No. 3;

(b) the amino acid sequence set forth in Seq. I.D. No. 4;

(c) amino acid sequences having at least 70% sequence identity with the amino acid sequence of (a) or (b); and

(d) amino acid sequences having at least 90% sequence identity with the amino acid sequence of (a) or (b).

14. A recombinant expression cassette comprising a promoter sequence operably linked to a nucleic acid molecule according to claim 13.

15. A transgenic plant comprising a recombinant expression cassette according to claim 14.

16. A recombinant nucleic acid molecule comprising a promoter sequence operably linked to a nucleic acid sequence, wherein the promoter sequence comprises a *SBP1* or *SBP2* promoter.

17. A recombinant nucleic acid molecule according to claim 16 wherein the promoter sequence comprises at least 25 consecutive nucleotides of a sequence selected from the group consisting of:

(a) Seq. I.D. No. 7; and

(b) Seq. I.D. No. 8.

18. A recombinant nucleic acid molecule according to claim 17 wherein the nucleic acid sequence encodes a plant sucrose binding protein.

19. A transgenic plant comprising a recombinant nucleic acid molecule according to claim 17.

20. A transgenic plant comprising a recombinant nucleic acid molecule according to claim 18.

```
sbp1 MAMRTKLSLA IFFFFLLALF SNLAFGKCKE TEVEEEDPEL     40
sbp2 MATRAKLSLA IFLFFLLALI SNLALGKLKE TEV.EEDPEL     39

sbp1 VTCKHQCQQQ QQYTEGDKRV CLQSCDRYHR MKQEREKQIQ     80
sbp2 VTCKHQCQQQ RQYTESDKRT CLQQCD...S MKQEREKQVE     76

sbp1 EETREKKEEE SREREEEQQE QHEEQDENPY IFEEDKDFET     120
sbp2 EETREKE...  ....EEHQEQ HEEEEDENPY VFEEDKDFST    109

                    *  *                 *         *
sbp1 RVETEGGRIR VLKKFTEKSK LLQGIENFRL AILEARAHTF     160
sbp2 RVETEGGSIR VLKKFTEKSK LLQGIENFRL AILEARAHTF     149
                   QR                              P

       *   *           *
sbp1 VSPRHFDSEV VFFNIKGRAV LGLVSESETE KITLEPGDMI     200
sbp2 VSPRHFDSEV VLFNIKGRAV LGLVRESETE KITLEPGDMI     189

         *   *                 *                 *
sbp1 HIPAGTPLYI VNRDENDKLF LAMLHIPVSV STPGKFEEFF     240
sbp2 HIPAGTPLYI VNRDENEKLL LAMLHIP..V STPGKFEEFF     227

                     **
sbp1 GPGGRDPESV LSAFSWNVLQ AALQTPKGKL EKLFDQQNEG     280
sbp2 GPGGRDPESV LSAFSWNVLQ AALQTPKGKL ERLFNQQNEG     267

              *
sbp1 SIFAISREQV RALAPTKKSS WWPFGGESKP QFNIFSKRPT     320
sbp2 SIFKISRERV RALAPTKKSS WWPFGGESKA QFNIFSKRPT     307

       *  *                                 *
sbp1 ISNGYGRLTE VGPDDDEKSW LQRLNLMLTF TNITQRSMST     360
sbp2 FSNGYGRLTE VGP.DDEKSW LQRLNLMLTF TNITQRSMST     346
                                               G
```

FIG. 1(a)

```
     *                   *
sbp1 IHYNSHATKI ALVIDGRGHL QISCPHMSSR SSHSKHDKSS  400
sbp2 IHYNSHATKI ALVMDGRGHL QISCPHMSSR SD.SKHDKSS  385
     P

                *          *          *
sbp1 PSYHRISSDL KPGMVFVVPP GHPFVTIASN KENLLMICFE  440
sbp2 PSYHRISADL KPGMVFVVPP GHPFVTIASN KENLLIICFE  425

          *        *                  * *
sbp1 VNARDNKKFT FAGKDNIVSS LDNVAKELAF NYPSEMVNGV  480
sbp2 VNVRDNKKFT FAGKDNIVSS LDNVAKELAF NYPSEMVNGV  465

sbp1 FLLQRFLERK LIGRLYHLPH KDRKESFFFP FELPREERGR  520
sbp2 .......... .......... SERKESLFFP FELPSEERGR  485

sbp1 RADA*     524
sbp2 RAVA*     489
```

FIG. 1(b)

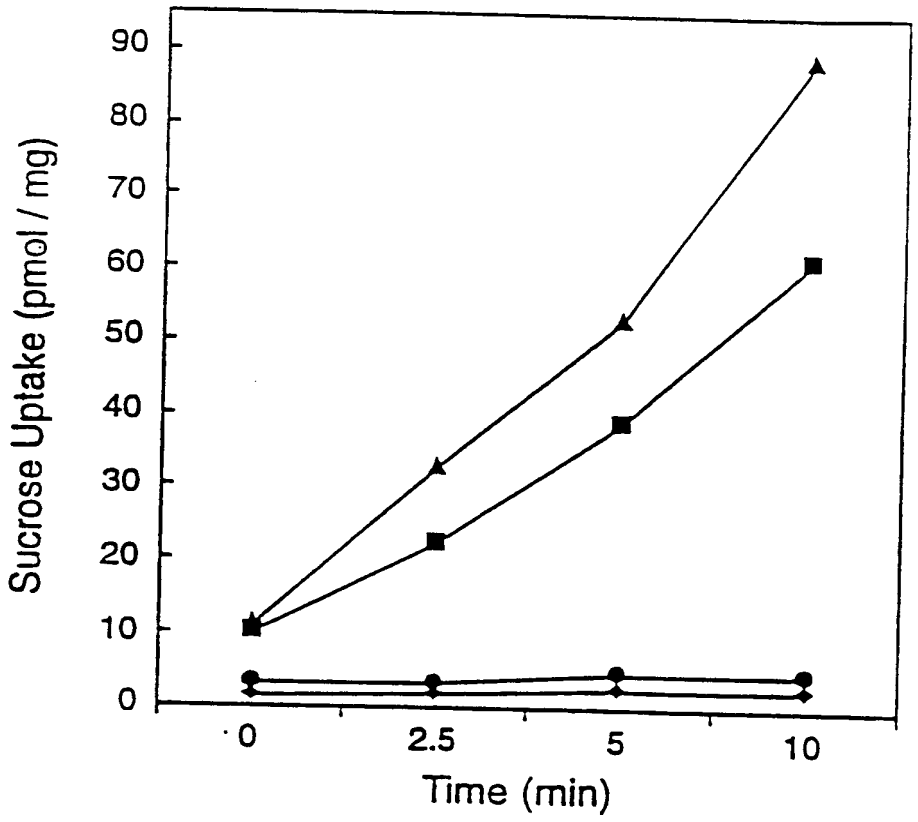

Sucrose Uptake (pmol / mg)
90
80
70
60
50
40
30
20
10
0
0
2.5
5
10
Time (min)

FIG. 2

INTERNATIONAL SEARCH REPORT

Inter ɔnal Application No
PCT/US 98/10465

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/82 C12N15/29 C12N1/19 C12Q1/68 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12Q A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	STOLZ, J., ET AL.: "rapid purification of a functionally active sucrose carrier from transgenic yeast using a bacterial biotin acceptor domain" FEBS LETTERS, vol. 377, 1995, pages 167-171, XP002079374 especially Fig. 3 see the whole document ---	1-3,8,9
A	WO 94 00574 A (INST GENBIOLOGISCHE FORSCHUNG ;FROMMER WOLF BERND (DE); RIESMEIER) 6 January 1994 page 4,5,9; page 17, line 10-14; examples, claims --- -/--	1-20

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 October 1998	13/10/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Holtorf, S

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	GRIMES, H.D., ET AL. : "a 62-kD sucrose binding protein is expressed and localized in tissues actively engaged in sucrose transport" THE PLANT CELL, vol. 4, December 1992, pages 1561-1574, XP002079375 cited in the application see the whole document ---	1-20
A	OVERVOORDE, P.J., ET AL. : "a soybean sucrose binding protein independently mediates nonsaturable sucrose uptake in yeast" THE PLANT CELL, vol. 8, February 1996, pages 271-280, XP002079376 cited in the application see the whole document ---	1-20
A	SAUER,N., ET AL. : "sugar transport across the plasma membranes of higher plants" PLANT MOLECULAR BIOLOGY, vol. 26, 1994, pages 1671-1679, XP002079377 see the whole document ---	1-20
A	SAUER, N., ET AL. : "suc1 and suc2: two sucrose transporters from arabidopsis thaliana; expression and characterization in baker's yeast and identification of the histidine-tagged protein" THE PLANT JOURNAL, vol. 6, no. 1, 1994, pages 67-77, XP002079378 ---	1-20
A	OVERVOORDE, P.J., ET AL. : "topological analysis of the plasma membrane-associated sucrose binding protein from soybean" THE JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 269, no. 21, May 1994, pages 15154-15161, XP002079379 cited in the application see the whole document ---	1-20
P,X	CHAO,W.S., ET AL. : "functional and structural analysis of sucrose binding protein in a heterologous yeast system" SUPPLEMENT TO PLANT PHYSIOLOGY, vol. 114, no. 3, July 1997, page 955 XP002079380 see the whole document --- -/--	1-9,11, 13,14, 16-18

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 98/10465

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	LU, M-Y. J., ET AL. : "site-directed mutagenesis of HIS71 in the proton-sucrose symporter" SUPPLEMENT TO PLANT PHYSIOLOGY, vol. 114, no. 3, July 1997, page 958 XP002079381 see the whole document	1,8,9

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 98/10465

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9400574 A	06-01-1994	DE 4220759 A	05-01-1994
		AU 671135 B	15-08-1996
		AU 4500393 A	24-01-1994
		CA 2137346 A	06-01-1994
		EP 0647273 A	12-04-1995
		HU 70472 A	30-10-1995
		JP 7509123 T	12-10-1995
		US 5608146 A	04-03-1997